

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 25, 2002, 02:34:30 ; Search time 6005.26 Seconds
(without alignments)
8999.927 Million cell updates/sec

Title: US-09-701-868-4
Perfect score: 2554
Sequence: 1 aaaaaagtttcaattttt.....aaaaaaaaaaaaaaaaaaaaa 2554

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0
Searched: 1797656 seqs, 10463268293 residues
Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:

- 1: gb_ba.*
- 2: gb_htg.*
- 3: gb_in.*
- 4: gb_om.*
- 5: gb_ov.*
- 6: gb_pat.*
- 7: gb_ph.*
- 8: gb_pl.*
- 9: gb_pr.*
- 10: gb_ro.*
- 11: gb_sts.*
- 12: gb_sy.*
- 13: gb_un.*
- 14: gb_vi.*
- 15: em_ba.*
- 16: em_fun.*
- 17: em_hum.*
- 18: em_in.*
- 19: em_mu.*
- 20: em_om.*
- 21: em_or.*
- 22: em_ov.*
- 23: em_pat.*
- 24: em_ph.*
- 25: em_pi.*
- 26: em_ro.*
- 27: em_sts.*
- 28: em_un.*
- 29: em_vi.*
- 30: em_htg_hum.*
- 31: em_htg_inv.*
- 32: em_htg_other.*
- 33: em_htgo_inv.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No. Query Score Match Length DB ID Description

1	2524.2	98.8	2653	8	LES012798	1
2	2524	98.8	2532	8	AF020390	Lycopersi
3	1842.6	72.1	2629	8	AY029226	Capsicum
4	1296	50.7	2753	8	AF064786	Carica pa
5	1265.8	49.6	2643	8	AB046543	Pyrus pyr
6	1256.2	49.2	2616	8	MAUBGALRP	L29451 Malus domes
7	1231.2	48.2	2783	8	CAAJ5042	Cicer ari
8	1218.8	47.7	2821	8	CAR011010	AJ011010 Cicer ari
9	1209	47.3	2166	8	AF229794	AF229794 Vigna rad
10	1207.4	47.3	2628	8	LAN011047	AJ011047 Lupinus a
11	1205.8	47.2	2628	6	A46218	Sequence 1
12	1205.8	47.2	2628	6	AR085663	Sequence
13	1182.4	46.3	2704	8	FRX278705	FR0278705 Fragaria
14	1181.6	46.3	2202	8	AF004812	AF004812 Mangifera
15	1172.8	45.9	3010	8	LES012796	AJ012796 Lycopersi
16	1172.8	45.9	3038	8	AF154421	AF154421 Lycopersi
17	1171.2	45.9	2944	6	A46235	Sequence 18
18	1171.2	45.9	2944	6	AR085670	Sequence
19	1168.2	45.7	2945	6	A44190	Sequence 2
20	1168.2	45.7	2945	6	AR029613	Sequence
21	1168.2	45.7	2945	8	LEBETAGG	X83854 L.esculentu
22	1168.2	45.7	2989	8	LES012797	AJ012797 Lycopersi
23	1168.2	45.7	3205	8	AF023847	AF023847 Lycopersi
24	1146	44.9	3309	8	FRX278703	FR0278703 Fragaria
25	1132.2	44.3	3283	8	ATH270297	ATH270297 Arabidops
26	1109.2	43.4	3124	8	AY064690	AY064690 Arabidops
27	1108	43.4	2235	8	AY064690	AY064690 Arabidops
28	1108	43.4	2471	8	ATH270300	ATH270300 Arabidops
29	1108	43.4	2504	8	AY054589	AY054589 Arabidops
30	1105.6	43.3	3241	8	AB061017	AB061017 Persea am
31	1099.6	43.1	2221	8	ATH270308	ATH270308 Arabidops
32	1088.4	42.6	2476	8	AF367327	AF367327 Arabidops
33	1082	42.4	2442	8	ATH270298	ATH270298 Arabidops
34	1076.6	42.2	3152	8	AOBGMAL	X77319 A. officinal
35	1004.8	39.3	2664	8	CAR012687	AJ012687 Cicer ari
36	986.4	38.6	2855	8	AY043231	AY043231 Vitis vin
37	958.4	37.5	2656	8	CAR6771	AJ006771 Cicer ari
38	952.4	37.3	2748	8	AY056285	AY056285 Arabidops
39	952.4	37.3	3067	8	ATH270299	ATH270299 Arabidops
40	936.8	36.7	2554	8	AF229795	AF229795 Vigna rad
41	934.2	36.6	3048	8	AF154424	AF154424 Lycopersi
42	924	36.2	2388	8	ATH270301	ATH270301 Arabidops
43	922.4	36.1	2451	8	AY069911	AY069911 Arabidops
44	920.8	36.1	2445	8	AY058098	AY058098 Arabidops
45	731.2	28.6	2781	8	ATH270304	ATH270304 Arabidops

ALIGNMENTS

RESULT 1
LES012798
LOCUS LES012798 2653 bp mRNA linear PLN 07-JAN-1999
DEFINITION Lycopersicon esculentum mRNA for ss-galactosidase, clone tEG3.
ACCESSION AJ012798
VERSION AJ012798.1 GI:4138140
KEYWORDS beta-galactosidase; ss-1,4-exogalactanase; ss-galactosidase; tEG3 gene.

SOURCE tomato.
ORGANISM Lycopersicon esculentum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I; Solanales; Solanaceae; Solanum; Lycopersicon.

REFERENCE 1 (bases 1 to 2653)

AUTHORS De Silva J.
TITLE Direct Submission
JOURNAL Submitted (11-NOV-1998) de Silva J., Plant Sciences, Unilever Research, Unilever Research, Colworth Laboratory, Sharnbrook, Bedford, MK44 1LQ, UK
FEATURES Location/Qualifiers
1. .2653 /organism="Lycopersicon esculentum"

/cultivar="Money maker"		
/db_xref="taxon:4081"		
/clone="tEG3"		
/dev_stage="Fruit"		
sig_peptide	91..159	
CDS	/gene="tEG3"	
	91..2265	
/gene="tEG3"		
/standard_name="beta galactosidase"		
/EC_number="3.2.1.23"		
/protein_id="CAAL0175.1"		
/function="ss-1,4-exogalactanase"		
/codon_start=1		
/evidence=experimental		
/product="ss-galactosidase"		
/db_xref="GI:4138141"		
/translation="MLRNVLLLVICLLDFSSVKASVSDRAIIINGRRKILISG SIHPRSPQMPDLIOKAKDGLDVIETVFWNGHGPSKYNFEGRYDLVRIKVV QRAGLYVNLRIPIYCAEWNFGFPVWLKYVPFMEFRNNQNPFFKVMHGFYVOKIVNMH KSENLFESQGGPIIMAOIENIEYGPVEWIEGAPKAYTKWAAQMAVLGTFPWIMCKQ EDAPDVIDCNQFYCEGFRPNKPYKPKMWTVEVMTGTYTFKFGPIQRPABDIAFSVA RFVONGSFFNVYVHGTFNFGRTSSGLFIATSYDYDAPLDEYLLNPEPKYGLHRLDH KAILSEPALVSSAAVTSLSQNOEAHVYRSKSGACAFLSNYDSRYSVKVTFFONRPY NLPSWISILPDKTAVTYNTAQVNSQSSIRKMTFAGGSLSQSNEETPTADDSGLTL ANGWEQKRVTRDSDYLWYMTNWNINIASNEIGFLKNGKDPYLTVMASGHLVHFVNGKL SGTVYGLDNPFLKYSYGNVKLRAGINKISLSLSVGLPNVGVHYDTHNAGVLGPTLS GLNCSRLNAQKWSYKVLKSGESLSLSLSSSSVWVRVGSILVAQKQPLTWYKATFN APGNDPLALDMSMGKQIWIINGEGVRHWPVYIAGDCSKSVYAGTFNEKKKQTC GOPSORWYHVPRLKPSGNLLVVFEEWGNPTGISLVRSR"		
91..2265		
gene	/gene="tEG3"	
160..2262		
/gene="tEG3"		
/EC_number="3.2.1.23"		
/function="ss-1,4-exogalactanase"		
/product="ss-galactosidase"		
BASE COUNT	861 a 441 c 575 g 776 t	
ORIGIN		
Query Match 98.8%; Score 2524.2; DB 8; Length 2653;		
Best Local Similarity 99.3%; Pred. No. 0;		
Matches 2535; Conservative 0; Mismatches 18; Indels 0; Gaps 0;		
Qy	1 aaaaaaagtccaatttttttctaaaaataaaaaaaattcatttttttgaaatggaa	60
Db	28 AAAAAAAGTTTCAATTTTTTCTRAAATAAAAAAATTCATTTTTTTGAATGGAA	87
Qy	61 aaaaagtcaagactaagtgtgtgtgtattagttattgtttattggattttttct	120
Db	88 AAAATGCTAAGGACTAATGTGTGTGTATTAGTTATTGTATTGTGATTTTTTTCT	147
Qy	121 tcagtgaaagctagtggtttcttatgacagagctataatcataaagggaagaaaa	180
Db	148 TCAGTGAAAGCTAGTGTTTCTTATGATGACAGAGCTATTAATCAATATGGGAAAGAAA	207
Qy	181 attcttatttctggttccaattcatatccaagaagpactcccacagatggcctgatct	240
Db	208 ATTCATTATTCTGGTTCATTATATCCAAAGACATCCACAGATGTGGCCTGATCTT	267
Qy	241 atacaaaagcctaagatggagccttagatgttattgaaacttatgtttctggaatgga	300
Db	268 ATACAAAAGGCTAAAGATGGAGGCTTAGATGTTATTGAAACTTATGTTTTCTGGAATGGA	327
Qy	301 catgagccttctcctggaaaaataaattttgaaagaagatatgattcttggattcatc	360
Db	328 CATGGCCCTTCCTGGAAAAATATAATTTGAAGGAAGATATGATCTTGTATGATTCATC	387
Qy	361 aaaaaggtacaaagacagagactttatgtcaattacgtattggcccttaactctgtgct	420
Db	388 AAAATGGTACAAAGACGAGACTTATGTCATTTACGTATTGGCCCTTACGCTCTGCT	447
Qy	421 gaatggaactttggggattccctgtttggctaaaaatatgtgcctggatggaattaga	480
Db	448 GAATGAAACTTTGGGGGATTCCTCTGTTTGGCTAAAAATATGTGCCCTGGTATGAAATTTAGA	507
Qy	481 acaaaaacaatcagccttttaagtggtctatgcaagatttgttccagaaaaatagccaatg	540
Db	508 ACAAAACATCAGCCCTTTAAGGTGGCTATGCGAGGATTTGTTTCAGAAAATAGTCAACATG	567
Qy	541 atgaagtccagaaaaattgtttgaaatctcaagaggagaccataaattatggcccagatacaa	600
Db	568 ATGAAGTCAGAAAAATTTGTTTGAATCTCAAGGAGGACCAATAATTTATGGCCCCAGATAGAA	627
Qy	601 aatgagtagtgaaccagtagaataaggaaaaatttggtgctcctggttaaaagcttatatacaaaatgg	660
Db	628 AATGAGTATGGACCAGTAGAATGGAAAAATGGTGTCTCTGGTAAAGCTTATACAAAAATGG	687
Qy	661 cgagctcaaatggctgtgtagtttgaaactggtgtcccatgatcatgtgtaagcaagag	720
Db	688 GCAGCTCAAAATGGCTGTAGTTTGAAACTGGTGTGCCCATGGATCATGTGTGAAGCAAGAG	747
Qy	721 gatgctcctgattcctgtgattgatacttgtaagtgttctactgcgaagggttccgctct	780
Db	748 GATGCTCCTGATCCTGTGATTGATTAATGCTTCTACTGCGAAGGGTTCCGTCTCT	807
Qy	781 aataagccttacaaacccataaagtgtggacagaagtatggactggtgtgtatacgaatttc	840
Db	808 AATAAGCCTTACAAACCTAAAAATGTGGACAGAAGTATGGACTGGCTGGTATACGAATTC	867
Qy	841 ggtgtccaattccotcaagaccgccgaagacattgcatcttcagttggtcaggtttgtt	900
Db	868 GGTGTCCCAATTCCTCAAGACCACCCGGAAGACATTTGCAATTTTCAGTTGCCAGGTTTGT	927
Qy	901 cagaaacaatggttcattcttcaattactacatgatatcatgtagggaggaaacaaattttggccgg	960
Db	928 CAGAACAAATGTTTCAATCTTCAATTTACTACATGTATCATGTGGAGGAACAAATTTTGCCCGG	987
Qy	961 acatcatcagggtcttctatgcaactagctacgattatgattgctcctcctcgatgaatat	1020
Db	988 ACATCATCAGGGCTTTTCAATTCACAACTAGCTACGATTAATGATGCTCTCTCGATGAATAT	1047
Qy	1021 gggttgctgaatgaaccacaagtatgggcacttgagagacttacataaaagctatcaagcta	1080
Db	1048 GGGTGTCTGAATGAACCAAGATATGGGCACCTTGACAGACTTACATAAAGCTATCAAGCTA	1107
Qy	1081 tctgaacgggttttagtttcatcatatgctgoggtgactagttcttgggaagtaatacaagag	1140
Db	1108 TCTGAACGGGCTTTAGTTTCAATCATATGCTGGGTGACTAGTCTTGGAGTAAATCAAGAG	1167
Qy	1141 gctaatgtttatagatacaaaatcctggagctgtgctgctttttttatccaactatgactct	1200
Db	1168 GCTCATGTTTATAGATCAAAATCTGGAGCTTGCTGTGCTTTTTTATCCAACATPATGACTCT	1227
Qy	1201 agatattcagtaaaagtacaccttcagaaataggccatacaatctgcctccatgggtccatc	1260
Db	1228 AGATATTTCAGTAAAAAGTCACTTTTCAGAAATAGGCCATCAATCTGCCCTCCATGGTCCATC	1287
Qy	1261 agcatcttcccagactgcaaaaactgccgttttacaaactgcacaggtttaactctcaaaagc	1320
Db	1288 AGCATTTCTCCGACTGCAAAACATGCCGTTTACAACTGCACAGGTTAACTCTCAAAGC	1347
Qy	1321 tcagacataaagatgacgcctgcaggttggtggtattctcttggcagctacatacaaatgaaga	1380
Db	1348 TCGAGCTAAAGATACGCGCTGCAGGTGGTGGATTGTCTTGGCAGCTACATAATGAAGAA	1407
Qy	1381 acgctactgctgtagacagagatacacttacagctaacgagactatggaaacagaaaaac	1440
Db	1408 AGCCTACTGCTGATGACAGCGATACATTACAGCTAACGGACTATGGGAACGAAAAAAC	1467
Qy	1441 gtcaacaagagattccatcagactatctgtgtgtacatgacaaaatgtaaaatagcatcctaat	1500
Db	1468 GTCAACAAGAGATTTCAGACTTATCTGTGGTACATGACAAATGTAAATATAGCATCTAAT	1527
Qy	1501 gaagatttctaaagaacggaaagatccttctcactgttatgtcctgctggttcgctggtc	1560

Db	1528	GAAGGATTTCTAAGAACGCGAAAGGATCCTTATCTATCTACTGTTATGTCGCGTGGTCATGTC	1587
Qy	1561	ttgcatgttttcgtaacatggaacacatcaggaactgtttatggtacattggataatacca	1620
Db	1588	TTGCATGTTTTCGTCATATGGAACACTATCAGAACTGTTTATGTTACATTGGATAATCCA	1647
Qy	1621	aaacttacatcagttggaacgtaagtttagagctgtgtatatacaagatttctctgtc	1680
Db	1648	AAACTTACATACAGTGGCAAGCTGAAGTTAAGAGCTGTTATTAACAAGATTCTCTGCTC	1707
Qy	1681	agtgtttccgtgtctctccaaacgttgcgtgcattatgatacatggaatgcaggagtt	1740
Db	1708	AGTGTTCCTGTTGCTCTCCGAACGTTGCGTGCATTATGATACATGAATGCAGGAGTT	1767
Qy	1741	ctaggtccagtcacgttgagcggtctcaatgaaggtcaagaacctgtgcgaacagaaaa	1800
Db	1768	CTAGTCCAGTCACGTTGAGCGGCTCAATGAGGGTCAAGAACTGGCGGAACAGAAA	1827
Qy	1801	tggcttacaaggtgtctgaaaggogaatcgtaagcttcttcaactccttaagtggaggt	1860
Db	1828	TGGTCTTACAAAGGTTGGTCTGAAAGCGCAATCGTTAAGTCTTCACTCCTTAAGTGGAGT	1887
Qy	1861	tctctgttgaatgggttcaggttccactaatggtcctcaaaagcagccctgaactgttac	1920
Db	1888	TCTTCTGTGAATGGTTTCGAGGTTCTACTAGTGGCTCAAAAGCAGCCCTGACTGGTAC	1947
Qy	1921	aaggtacatttaacgcgcctggaggaatgatccactagcttttagacatggcaagtatg	1980
Db	1948	AAGGCTACATTTAAACGCGCTGGAGGAATGATCCACTAGCTTTAGACATGGCAAGTATG	2007
Qy	1981	ggaaaaggtcagatatggaataatggtaagcgtagtgcgcctattggcctggatataca	2040
Db	2008	GCAAAAGGTCAGATGATGATAATGGTGAAGCGTAGTGCCTCCATTTGGCCTGGATACATA	2067
Qy	2041	gcacaagcgactcagcaaatcagttatctgtaagcttgaagcttcaacagagaagaagtgcga	2100
Db	2068	GCAACGCGACTCAGCAAAATGAGTTATGCTGGAAGCTTCAACGAGAAAGTGGCCAG	2127
Qy	2101	actaacgtcggaacacaccttcagagatggtaccatgttccacgatcgtggctgaaccca	2160
Db	2128	ACTAACTGCGGACAACTTCTCAGAGATGTTACCATGTTCACGATCGTGGCTGAACCA	2187
Qy	2161	agtggaactgttagtagtattcgaagaaatggggaggttaataccaacaggaatttctcta	2220
Db	2188	AGTGGAACTTGTAGTAGTATTCGAAGAATGGGAGGTAATCCAACAGGAATTTCTCTA	2247
Qy	2221	gtcaggagatcaagataaagaactcgaagaatgaactgttccagtaactatggtgcttg	2280
Db	2248	GTCAGGAGATCAAGATAAAGAACTCGAAAAGTAAACTTGTTCAGTAACATATGCTGCTG	2307
Qy	2281	aattcgcgcgcaaaaaacacacacgaagctcaaatggaggtcacagtttgcaaatgc	2340
Db	2308	AATTGCGCGCAAAAATACATACACGAAGCTTAACAATGGAGGCTACAGTTTGCAAAATGC	2367
Qy	2341	agctgaataaacattagaagataaagaataatttattgataaaaagagtatataaatttac	2400
Db	2368	AGCTGAATAAACAATAGAGATAAAGAAATATTGTATTAAGAGAGTATATAAATTATC	2427
Qy	2401	agagaatttcttattcttctgtaaaacttggtttataaagtttatcaagaattttctg	2460
Db	2428	AGAGAATTTCTTATTCTTTGTAACACTTTGGTTTATAAAGTTTATACAGAATTTCTG	2487
Qy	2461	ttatttgattatgagattgagaagattgtcacagcttcccaataactattagatacaaaa	2520
Db	2488	TTATTGATATATGAGATTGAAGAATTGTACAGCTTCCAAATACTATTAGATACAAA	2547
Qy	2521	taatttctaatgtaaaaaaataaaaaaataaaaaa 2553	
Db	2548	TAAATTTTCATGTAACATATATACAAATTTATACAA 2580	
RESULT 2			
AF020390			

LOCUS	AF020390	2532 bp	mRNA	linear	PLN 25-JUL-2000
DEFINITION	Lycopersicon esculentum beta-galactosidase (TBG4) mRNA, complete cds.				
ACCESSION	AF020390				
VERSION	AF020390.2				
KEYWORDS	GI:7923928				
SOURCE	tomato.				
ORGANISM	Lycopersicon esculentum				
REFERENCE	Smith, D.L., Starr, D.A. and Gross, K.C. A gene coding for tomato fruit beta-galactosidase II is expressed during fruit ripening. Cloning, characterization, and expression pattern				
AUTHORS	Smith, D.L., Starr, D.A. and Gross, K.C.				
TITLE	A family of at least seven beta-galactosidase genes is expressed during tomato fruit development				
JOURNAL	Plant Physiol. 117 (2), 417-423 (1998)				
MEDLINE	98289087				
PUBMED	9625694				
REFERENCE	2 (bases 1 to 2532)				
AUTHORS	Smith, D.L. and Gross, K.C.				
TITLE	A family of at least seven beta-galactosidase genes is expressed during tomato fruit development				
JOURNAL	Plant Physiol. 123 (3), 1173-1184 (2000)				
MEDLINE	20349728				
REFERENCE	3 (bases 1 to 2532)				
AUTHORS	Smith, D.L., Starr, D. and Gross, K.C.				
TITLE	Direct Submission				
JOURNAL	Submitted (21-AUG-1997) HCQL, USDA/ARS, 10300 Baltimore Ave., Beltsville, MD 20705, USA				
REFERENCE	4 (bases 1 to 2532)				
AUTHORS	Smith, D.L., Starr, D. and Gross, K.C.				
TITLE	Direct Submission				
JOURNAL	Submitted (18-MAY-2000) HCQL, USDA/ARS, 10300 Baltimore Ave., Beltsville, MD 20705, USA				
REMARK	Sequence update by submitter				
COMMENT	On May 18, 2000 this sequence version replaced gi:3299895.				
FEATURES	Location/Qualifiers				
source	1..2532				
gene	/organism="Lycopersicon esculentum"				
sig_peptide	/cultivar="Rutgers"				
CDS	/db_xref="taxon:4081"				
	/chromosome="12"				
	/map="R1 line; overlap of IL 12-2, IL 12-3"				
	/tissue_type="fruit"				
	1..2532				
	/gene="TBG4"				
	64..132				
	/gene="TBG4"				
	64..2238				
	/gene="TBG4"				
	/function="beta-galactosidase activity as measured by release of galactose from synthetic substrates such as p-nitrophenyl-beta-D-galactopyranoside; exo-galactanase activity as measured by release of galactose from a variety of native cell wall substrates"				
	/note="exo-galactanase; secreted protein; can release galactose from lactose"				
	/codon_start=1				
	/product="beta-galactosidase"				
	/protein_id="AAC25984.1"				
	/db_xref="GI:3299896"				
	/translation="MLRTNVLLVLLVLCIDFFSSVKASVSDDDRAILLKRLKILISG SIHPPSTPQWFDLTKQKADGGDLVITVFNHGPSPKYNFGRDVLRFKMY QRAGLVNLRIGYVCAEMNFGFPVWLKYVPGNEFTNNQPKVAMQGVQKIVNM KENLFESQGGPIIMAEIENEYGEVEIGAPKAYTKWAAQAVGLKGTGPVIMCKQ EDAPDPVIDTCNGFYCEGPNKPKMTEVMTGWTFFGGPIQORPAEDIAFVA RFVONSGFPNYVMHGTNFGTSSGLFATSYDADLDEYGLLNEPKYGHRLRH KAIKLEPALVSSYAATVSLGSNQAHVFSKSCACAAFLSNYDSYKVTQNPY NLFPWSISLDPCKTAVNTAQNSQSSSMTAGGLSNQSWSYNETPADDSDLT ANGLEQKNVTRDSSDYLWMTNNTASNEGFLKNGKPYLTVMASGHLHVFVNGKL SGTVYGTLDNPKLTSGNKLKRLAGINKISLLSVSLVGLPNVGVHYDTWNAGVLGPVTL				

GLNBSRLAKQKSYKVGLKESLSLSLSGSSSVVWRGSLMAQKPLTWYKATFN
GPGNDPLALDMSMGKQIWIWEGVGRHWPYIAQGDCKSKCYAGTFNBKKQTNIC
QPSQWRWHPRLSRSLKPSGNLLVFEWGGNPTGISLVRSR*
BASE COUNT 800 a 431 c 559 g 742 t
ORIGIN

Query Match 98.8%; Score 2524; DB 8; Length 2532;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 2527; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 aaaaaaagtccaatttttttttctaaaaataaaaaaaatcoatttttttttgaaatgtaa 60
DB 1 AAAAAAAGTTTCAATTTTTTTCTAAAAATAAAAAAATTCATTTTTTTTGAATGGAA 60

QY 61 aaaaagtcaagactaagtgtgtgtgtattattgtttattgtttattgtttttttttt 120
DB 61 AAAATGCTAAGGACTAATGTGTGTGTATTAGTTATTGTGATTTTTTTTTTCT 120

QY 121 tcagtgaagctagtggtttctttatgatcacagagctataataatgggaaagaaaa 180
DB 121 TCAGTGAAGGCTAGTGTCTTTATGATGACAGAGCTATAATCATAAATGGGAAAGAAA 180

QY 181 attottatttctggttcaattcaattatccaagaagcactccaagatgtggcctgatctt 240
DB 181 ATTCTTATTTCTGGTTCAATTCATTATCCAAAGACCTCCACAGATGTGGCCTGATCTT 240

QY 241 atacaaaagcctaaagatggagccttagtgttattgaaactattgtttcttggaatgga 300
DB 241 ATACAAAGGCTAAGATGAGGAGCTTAGATGTATTGAAACTTATGTTCGGAATGGA 300

QY 301 catgagccttctcctggaaaaataattttgaaggaagatatgatctttagattcatc 360
DB 301 CATGAGCCTTCCTCGGAAAAATAAATTTGAAGGAAGATATGATCTTGTTAGATTTCATC 360

QY 361 aaaaagggtacaaagacgaggaacttatgtcoaatttcaatttggccttaagctgtgct 420
DB 361 AAAATGGTACAAAGACAGGAGCTTATGTCAATTTACGTATTGGCCCTTACGCTGTGCT 420

QY 421 gaatggaactttgggggttccctgtttggctaaaaatgtgcctgggtatggaatttga 480
DB 421 GAATGGAACCTTTGGGGGATTCCTGTGTGGCTAANAATATGTGCCCTGTGTAATTTAGA 480

QY 481 acaacaatcagccctttaagtggtgctatgcaaggaatttggtcagaaataagtcacatg 540
DB 481 ACAACAATCAGCCTTTAAAGGTGGCTATGCAAGGATTTGTTCAAGAAAAATAGTCAACATG 540

QY 541 atgaagtcagaaaaatttggtaattctcaaggaaggaccaaataattatggccccagatacaa 600
DB 541 ATGAAGTCAGAAAAATTTGTTGAAATCTCAAGGAGGACCAATAAATATGGCCAGATAGAA 600

QY 601 aatgagatggaccagtagaattgggaaattggtgctcctggttaaaagcttatacaaaatgg 660
DB 601 AATGAGTATGGACCAAGTAGAATGGGAATTTGGTCTCTGTTAAAGCTTATACAAAAATGG 660

QY 661 gcagctcaaatggctgtaggtttgaaactgggtgtcccatggaatcatgtgtaagcaagag 720
DB 661 GCAGCTCAAAATGGCTGTAGGTTGAAAACTGGTGTCCCATGGATCATGTGTAAGCAAGAG 720

QY 721 gatgctcctgatcctgtgatgatgatacttgtaattggtcttactgaggaaggttccgtcct 780
DB 721 GATGCTCCTGATCCGTGATTGATGATCTTGAATGGCTTCTACTGGGAAGGGTTCCGTCTCT 780

QY 781 aataagccttacaacaccttaaaatgtggacagaagtatggactggcgtggtatcacgaaattc 840
DB 781 AATAAGCCTTACAAACCTTAAATGTGGACAGAGATATGGACTGGCTGATACGAAATTC 840

QY 841 ggtgtccaattcctcaagaaccagccgaagacatatgcaatttcaagttgaggtttggtt 900
DB 841 GGTGTCCAATTTCCCTCAAGACACAGCCGAAGACATTTGCAATTTTCCAGTTGCCAGGTTTGT 900

QY 901 cagaacaatggttcaatttcaattactacatgatatcatggaggaacaaatttggccgg 960
DB 901 CAGAACAAATGTTCAATTTCTTCAATTTACTACATGTTATCTGAGGAGAACAAATTTTGGCCGG 960

DB 901 CAGAACAAATGTTCAATTTCTTCAATTTACTACATGTTATCTGAGGAGAACAAATTTTGGCCGG 960

QY 961 acatcatcagggcttttcttattgaaactagctacagattatgatgtcctctcctgatgaat 1020
DB 961 ACATCATCAGGGCTTTTTCATTGCAACTAGCTACGATTTATGATGCTCTCTCGATGAATAT 1020

QY 1021 ggggttgctgaatgaaccaaaagtatgggcacttgagagacttacataaaagcttatcaagcta 1080
DB 1021 GGGTTCGTGAATGAACCAAGATATGGCACCTTGAGAGACTTACATAAAGCTATCAAGCTA 1080

QY 1081 tctgaacccggcttttagttttcatcatatgctgcgggtgacttagctcttggaagtaatacaagag 1140
DB 1081 TCTGAACCCGGCTTTAGTTTTCATCATATGCTGCGGTGACTAGTCTTTGGAAGTAAATCAAGAG 1140

QY 1141 gctcatgtttatagatcaaaaactctggagctgtgctgtcttttttttccaaactatgactct 1200
DB 1141 GCTCATGTTTATAGATCAAAATCTGGAGCTTGTGCTGCTTTTATCCAACTATGACTCT 1200

QY 1201 agatattcagtaaaagtccacttttcagaaataggccatacaaatctgcctccatggctccatc 1260
DB 1201 AGATATTTCAGTAAAAAGTCACCTTTTCAGAAATAGGCCATACAAATCTGCCCTCCATGGTCCATC 1260

QY 1261 agcattcttcccgactgcaaaaactccgcttttaacaactgcacaggttaactctcaagac 1320
DB 1261 AGCATTTCTCCGACTGCAAAACTGCCCTTTTACAACTGCACAGGTTAACTTCTCAAGAC 1320

QY 1321 tcgagcataaagatgacccctgcagctgtgtggtattctcttgccagctacatacaaatgaagaa 1380
DB 1321 TCGACATTAAGATGACCCCTGCAGGTGGTGGATTGCTCTTGGCAGCTATACATGAAGAA 1380

QY 1381 acgctactgtgatlgaacagcgatacacttacagctaacaggactatgggaacagaaaaac 1440
DB 1381 ACGCCTACTGTGATGACAGCGATACACTTACAGCTAACGGACTATGGGAACAGAAAAAC 1440

QY 1441 gtcacaagagattcatcagactatctgtggtacatgcacaaatgtaaatatagcatcatt 1500
DB 1441 GTCAAGAGAGATTTCATCAGACTATCTGTGGTACATGACAAATGTAAATATAGCATCTAAT 1500

QY 1501 gaagattctcaagaacggaaagatccttatctcactgttatgtcgcgtggtcatgct 1560
DB 1501 GAAGGATTTCAAGAACGGAAAGGATCCTTATCTCAGCTGTTATGTCGCTGGTCAATGTC 1560

QY 1561 ttgcatgttttctgctcaatggaaaaacta tcaggaactgtttatggtacatttggaatacca 1620
DB 1561 TTGCATGTTTTCGTCATATGGAAAACTATCAGGAACGTGTTATGGTACATTTGGATTAATCCA 1620

QY 1621 aaactacatcagtggaacgtaagtttaagagctgtatttaacaagatttctctgctc 1680
DB 1621 AAACCTACATCAGTGGCAACGTGAAGTTAAGAGCTGGTATTAAACAGATTCTCTGCTC 1680

QY 1681 agtgttccgttggctcccgaaacgttggtgcattatgatatacaggaatgcagagatt 1740
DB 1681 AGTGTTCGTTGGTCTCCCGAACGTTGGCGTGCAATTTATGATACATGGAATGCAGAGTT 1740

QY 1741 ctaggctcagtcacgtttgagcgggtctcaatgaagggttcaagaaacttggcgaacagaaa 1800
DB 1741 CTAGGTCAGCTCACGTTGAGCGGTCTCAATGAAGGGTCAAGAAACTTGGCGAAACAGAAA 1800

QY 1801 tggcttacaaggttggtctcaaaagcgaaatcgtttaagcttctcactccttaagtggagt 1860
DB 1801 TGGCTTTACAAGGTGGTCTGAAAGCGGAATCGTTAAGTCTTCACTCTCTTAAAGTGGGAGT 1860

QY 1861 tctctgttggaattgggttcgaggttccactaaatggctcaaaagcgccttacttgggtac 1920
DB 1861 TCTTCTGTTGAATGGTTTCGAGGTTTCACTAATGGCTCAAAAGACGCCCTCTGACTTGGTAC 1920

QY 1921 aaggtacatttaacgcgccttggaagaaatgatccacttagcttttagacatggcgaagtatg 1980
DB 1921 AAGGTACATTTAAGCGCCTTGGAAGAAATGATCCACTAGCTTTAGACATGGCAAGTATG 1980

QY 1981 ggaagagctcagatatgataaaatggtgaagcgttagtcgctcattgcccattggtgatacata 2040
DB 1981 GGAAGAGCTCAGATTGGATTAATGGTGAAGCGGTAGTGTGCCCATTTGGCTTGGATACATA 2040

Qy 772 ttccgctcctaataagccttacaacctaataatgtggacagaagtatgactggtcgttat 831
Db 721 TTCCGCTCTAATAAGCCTTACAAACCGAAATGTGGACAGAGTTTGGACGTGGATT 780
Qy 832 acgaaattcgttggttccaatctctcaagaccagccgagacattgcatctttcagttgcc 891
Db 781 ACAAAATTTGGCGGTCCAAATTTCTCAACGGCCAGCTGAAGATATCGCGTTTTCAGTTGCA 840
Qy 892 aggttggttcagaaacaaatggttcattcttctcaattactacatgatatcatgggaaacaaat 951
Db 841 AGGTGTTGTCAGAAACAGCGGTTTCACTTCAATTTACTACATGATACACGGGGGAAACAAAT 900
Qy 952 ttgtgcccagacatcatcaggggtttttcattgcaactagctacgattatgagtctctctc 1011
Db 901 TTTGCCGGGACATCTTCAGGGGCTTTTCATCGGACTAGCTATGATATGATGCTCCTATC 960
Qy 1012 gatgaatatgggttgcgaatgaacaaagatggtggcacttgagagattacataaagct 1071
Db 961 GATGAGTATGGGCTTTCTGAATGAACCAAGTACGGGCACCTGAGAGAGTTGCATAAAGCG 1020
Qy 1072 atcaagctatctgaaccggcttttagtttcaatcatatgctgcggtgaactagtttggaagt 1131
Db 1021 ATCAAGCAATGTGAACCGGCTTTAGTTTCATGCTATCCACGGTGACTAGTCTTGGTAGT 1080
Qy 1132 aatcaagaggtcattgtttatagatcaaaactctggagcttctgctgctttttttatccaac 1191
Db 1081 AATCAAGAGGCTCATGCTATAGATCGAAATCTGGAGCTTGTGCTGCAATCTCTGTCCAAAC 1140
Qy 1192 tatgactctagatattcagtaaaagtcaccccttcagaataggaatgcaaatctgcctcca 1251
Db 1141 TATGATGCTAAATATTCAGTTCGAGCTCAGCTTTTCAGAAATTTGCCATACGATCTTCCCTCG 1200
Qy 1252 tggctcatcagcattctccgactgcaaaactgcggttttcaaacactgcacaggttaac 1311
Db 1201 TGGTCCATCAGCATTTCTCCCGACTGCAAAACCTGTGTATTAAACACCCGAAAGGTTAGC 1260
Qy 1312 tctcaagctcagcataaagatgacgctgcaggtggtggtattctctggcagtcatac 1371
Db 1261 TCCCAAGGCTCGAGCATAAAGATGACGCTGCAGGTGGTGGATTGCTTGGCAGTCATAC 1320
Qy 1372 aatgaagaaacgctactcgtgtagacagcgatatacacttacagctaaacgactatggaa 1431
Db 1321 AATGAAGACACACCGACTGCTGATGACAGCGATACACTTAGAGCTTAACGGGACTATGGAA 1380
Qy 1432 cagaaaaacgtcacagaagattcatcacactatctgtgtgtagcatgacaaaatgtaaatata 1491
Db 1381 CAGAAACGTCACAGAGATTTCAATCGGACTATCTGTGTATACACAGCGTAAACATA 1440
Qy 1492 gcataatgaaggatttctaaagacggaaggtatctctcactgtttatgtccgct 1551
Db 1441 GCATCTAACGAAGGATTTCTAAGAGCGGAAGGATCTTATCTCACGGTTATGTCGGCT 1500
Qy 1552 ggtcatgttctgcatgttttctgctcaatggaacactatcaggaactgtttatggtacattg 1611
Db 1501 GGTCAATGTTGTCATGTTTCTGCTAATGGAACCTTGCAGGAACCTGATACGGGGCATTTG 1560
Qy 1612 gataatcaaaacttacatacagtggaacgtgaagtttaagagctggtattatacaagatt 1671
Db 1561 GATAATCAAAACTTACATACAGCGGCAATGTGAAGCTTAATGCGCGGTATAAACAAAGATT 1620
Qy 1672 tctctgctcaggttctcgttggtctcccgaaactggtggtggtcattatgatacattggaat 1731
Db 1621 TCTCTCTCAGTGTTCGGTTCGCCGACGCTTGGCGGTACATATGATGATGCTGGGAT 1680
Qy 1732 gcaggagttctaggtccagtcacgttgtagcggtctcaatgaaggtcgaagaacttggcg 1791
Db 1681 GCAGGAGTTCTAGTCCAGTCACGTTGAGCGGCTCTCAATGAGGGGTACAGAGACTTGGCG 1740
Qy 1792 aaacagaatggtcttacaaggttggtctgaagcggaatgtaagtcttcaactcctta 1851
Db 1741 AANCAAGAAATGGTCTATAAAGGTTGGTCTAAAGGCGGAATCGCTAAGTCTTTCACACGTTA 1800

Qy 1852 agtggaggttcttctgttgaaatggtgtcgaggttctactaaatggtcctaaagcagccctg 1911
Db 1801 AGTGTAGTCTTCTCTGTCGAATGGGTTCAAAGGTTCAAGGTTCACTGGTTCGAACACACCCCTG 1860
Qy 1912 acttggtcacaggctcacatttaacgcgctggagagaatgatccactagcttttagacatg 1971
Db 1861 ACATGTTACAAGGCAACATTTAGCGCGCTGGAGGAAATGAACCCGTAGCTTTAGACATG 1920
Qy 1972 gcaagtatggaaaaaggtcagatatgataaattggtgaagcgttagtgcgccattggcct 2031
Db 1921 GCGAGTATGGGAAAGGTCAGATATGGATCAATGTGAAGCGTAGGTCGCCATTGGCT 1980
Qy 2032 ggatacatagcacaaaggcagctgcagcaaatgacgttatgtctgggaacttcaacacagaag 2091
Db 1981 GGATACGCGACACAAAGCGACTGCAGCAAAATGCAGTTACGCTGGAAACATTCAACGAGAAG 2040
Qy 2092 aagtgcgaactaaactcgcggcaacctctctcagagatggtacacatggttccacgactcgtg 2151
Db 2041 AAATGTCAAGCTAATCTGCGGACAGCTTCTCAGAGATGATCATGTCCCGCATCGTGG 2100
Qy 2152 ctgaaaccaagtggaaacttcttagtatttcgaagaatgggaggttaataccaacagga 2211
Db 2101 CTGAAAACAAGTGGAAACTTGTGTAGTATTTGAAGAAATGGGTGGTGCATCCCAACAGA 2160
Qy 2212 atttctcagtcagagatcaagaataaagaaactcgaaagcttctcagtaactgttccagtaacta 2271
Db 2161 ATTTCTTTAGTAAGAAGATCAAGATTAAGAGCTTCAAAAGT----- 2201
Qy 2272 tgggtgctgaattcgcgcgaaaaatacatcacagaactaaacataatgaggtgaggtcacagttt 2331
Db 2202 -----AAAAATGCATACAGAAGCTAACAAATGAGGCGCTATA-TTT 2240
Qy 2332 gcaaatgcagctgaataaaaacttagaagataaagaataattgtattaaaggagata 2391
Db 2241 GCAAAATGCAGCTGAATAATCAATTGAG-TACATAAATAATTGATAAAGGTATATAA 2299
Qy 2392 taaat 2396
Db 2300 AAAAT 2304

RESULT 4
AF064786 2753 bp mRNA linear PLN 13-NOV-1998
LOCUS Carica papaya beta-galactosidase precursor, mRNA, complete cds.
DEFINITION
ACCESSION AF064786
VERSION AF064786.1 GI:3869279
KEYWORDS
SOURCE papaya.
ORGANISM Carica papaya
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eudicots II; Brassicales; Caricaceae; Carica.
REFERENCE
1 (bases 1 to 2753)
AUTHORS Othman, R., Choo, T.S., Ali, Z.M., Zainal, Z. and Lazan, H.
TITLE A full-length beta-galactosidase cDNA sequence from ripening papaya
JOURNAL plant. Physiol. 118 (3), 1102 (1998)
REFERENCE
2 (bases 1 to 2753)
AUTHORS Othman, R., Choo, T.S., Ali, Z.M. and Lazan, H.
TITLE Direct Submission
JOURNAL Submitted (12-MAY-1998) Biochemistry, Universiti Kebangsaan
Malaysia, Bangi, Selangor 43600, Malaysia
FEATURES
Location/Qualifiers
1..2753
/organism="Carica papaya"
/cultivar="Eksotika"
/db_xref="taxon:3649"
/clone="pPBG1"
/tissue_type="mesocarp"
/dev_stage="ripening fruit"
1..126
5'UTR 127..180
sig_peptide 127..2292
CDS

```

/EC_number="3.2.1.23"
/function="involved in cell wall depolymerization"
/codon_start=1
/product="beta-galactosidase precursor"
/protein_id="AAC77377.1"
/db_xref="GI:3859280"
/translation="MUKTNLIVLFLFCSLWLSVEAVSYDHHKAIILNRRRIILSGSI
HYPRSPQMPDLIQNAKEGLDVIQTYFWNGHEPSPGNYFYEDRYDLVKFLKLHQ
AGLYVHLRISPIICGWNFGFFWLKYVPGIQFRTDNGFPFKAQMKQFTEKIVMMKA
EKLFEPQGGPIINSQENEXYPIEWEIGAPGRKAYTKWAAQMAVGLGTGVPWIMCKQED
APDPIIDQNGFYENFMPNANYKPKMFTFAWTGWTEFGGPPVYRPAEDMAYSVARF
IQNRGSFINYMYHGTNFGRTAGPPIATSYDYDAPLDEYGLRREPCKWHLRDLHKT
IKCFESLVSVDPKVTSLSGNQEAHFVTKISCAAFLLANDLKYSVKVTQNLPLDLP
PWSVLSLPDCKTVFNTAKVYSQSLAKMIAVNSAFWSQSYNEETPSYDAVFTKDG
LWEQISVTRDQATLYMYMTDVTICPDEAFILKNGQDPILITVMSAGHALHVPVNGLSGT
VYQLENPKLAFSGKVLKRAVNVKLSLLSTAVGLPNVGLHFVLPVGLVGLVPLKGVN
SGTWDMSKKWSYKIGLKGPAALSHTVSSGSSVVEVGSLLAORQPLIWKYKTFNAPV
GNPDLADMNSKMGQIOWINGQSIGRHWFPYKARGSCGCACNYAGIIDEKKCHSNCGA
SQRWYHVPKSWLNPNTANLLVVFEEWGDPTKISLVKRVV"
mat_peptide
181..2289
/product="beta galactosidase"
3'UTR 837 a 504 c 607 g 805 t
BASE COUNT
ORIGIN

Query Match 50.7%; Score 1296; DB 8; Length 2753;
Best Local Similarity 74.9%; Pred. No. 2.1e-240;
Matches 1652; Conservative 0; Mismatches 545; Indels 9; Gaps 2;

QY 24 taaataaaaaaaattcatttttttgaatgtggaataaagtctaaggactagtgttt 83
DB 87 TGAAGGAGTAACAATATTATTGATTTTTCATTTGCTATCATGTGTAAGACAAAACCTGGT 146

QY 84 gtgttattagtattgtttattggaatttttttttttcagtgaaagctagtgttttta 143
DB 147 CTGTGCTTGTGTGTTGTCATGG-----CTTTGTGCTGTGTTGAAGCTACTGTGCTTTA 200

QY 144 tgatgacagagctataatcaataaagtggaaagaaaaattcttattctgtgtcaattca 203
DB 201 CGACATAAAGCTATAATCATTAATGCGCGCAGAAGGATCTTATTCTGGCTCCATTCA 260

QY 204 ttatccaagaagactccacagatgtgtgacctgattcttatcaaaaaggctaaaagtggag 263
DB 261 TTAATCCAGAAAGCACTCCTCAGATGTGGCCTGATCTTATACAAAATGCTAAAGAAGGAG 320

QY 264 cttagatgttatgaaactatgtttctggaatggacatgagccttctctggaataa 323
DB 321 GTTAGATGTCATACAGACTTATGTTTGTGGAACGGACATGAGCCCTCTCTCGAAATTA 380

QY 324 taattttgaaggagatatgatctgttagattcatcaaaatggatcaaaagagcaggact 383
DB 381 TTAATTTGAAGACAGGTATGATCTTTGATAAGTTTCATCAAGTTGTCATCAAGCTGGTCT 440

QY 384 ttatgtcaattacgtattgacctacgtctgtgtgctgaaatggaaactttggggattccc 443
DB 441 GTATGTTTCATCTCAGAATAAGTCCTTATATTGTGTGGTGAATTTGGGGGTTTTTC 500

QY 444 tgttctgataaataatgtcgtgattgaaatttagaacaacaatacagccttttaagg 503
DB 501 TGTGTGCTCAATACGCTTCTGTGATTTCAATTACAGACAGACAAATGGACCTTTCAAGGC 560

QY 504 ggcctatcaaggattttgttcagaaaaagtagcaacatgatgaagtcaagaaattgtttga 563
DB 561 ACAAATGCAAAAATTTACAGAGAAAATAGTCAACATGATGAAGCGCAGAAAAGTTATTGA 620

QY 564 atctcaaggaggaccaataattatggcccgatatacaaaatgagtatggaccagtagaattg 623
DB 621 ACCTCAAGGGGGTCCCAATAATTATGTACACATAGAGAATGAGTATGGACCTATTGAGTG 680

QY 624 ggaattgggtgctcctggttaagacttatcaaaaatggcagcctcaaatggctgtagttt 683
DB 681 GGAATTTGGAGCACCGGGGAAGCTTTATACAAAATGGGACACAAAATGGCAGTGGGTCT 740

```

```

QY 684 gaaactgggtcccatggtatcgtgttaaagaagagatgctcctgatcctgtgattga 743
DB 741 TGGCAGTGGAGTCCCATGGATTATGTGCAAGCAAGAGAGATGCTCTGACCCCAATTATTGA 800

QY 744 tacttgaatggcttctactcgcgaaggggtccgttccctaataagccttacaacacataa 803
DB 801 CACTTGAACATGGTTTCTATTGTGAAAAATTTTCATGCCAACGCCAACATACAAACCAAAA 860

QY 804 gtgacagaagatggactgctgtatatacaaaaattcgggtggtccaatctcctcaaaagc 863
DB 861 GTTTACAGAGCCCTGGAGTGGCTGGTACACGGAATTTGGCGGTCCAGTTCCTTTTAGACC 920

QY 864 agcgaagacattgcattttcagttgcaggtgttcttcagacaacaaatgttctattctcaa 923
DB 921 TGCAGAAGACATGGCTTACTCCGTTGCAAGGTTCATTTCAGAATAGGGGATCATTCATTAA 980

QY 924 ttactaatgtatcatggaggaacaaaattttggcggagacatcaagggcttttcattgc 983
DB 981 TTATTATATGTACCATGGAGGAACAAAATTTTGGCAGAACTGCTGGAGGTCTCTTCATTGC 1040

QY 984 aactagctacgattatgatcctcctcgtatgaatagatgggtgctgaatgaacaaagta 1043
DB 1041 TACTAGCTATGATTACGATGCCCTCTTGATGAGTATGGACTAAGGAGGGAGCCAAAATG 1100

QY 1044 tgggcacttgagagacttacataaagctatcaagctatctgaacggcgttttagtttcac 1103
DB 1101 GGGCAGCTTGAAGGATTTGTCATAAAACCATCAAAATTAATGTGAACCATCTTTTAGTTCTGT 1160

QY 1104 atagtgcgggtgactagcttggaagtaatacaagagctcatgtttatagatacaaaatc 1163
DB 1161 TGATCTTAAAGTGACATCGTTTAGGAAGTAACCAAGAGGCTCATGTGTTTGGACAAAAC 1220

QY 1164 tggagcttgctcctctttttatccaactatgactctatgatattcagtaaaagtcac 1223
DB 1221 ---CTCTTGTGCTGCATTCCTTGTGTAACCTAGATCTGAAGTACTCAGTAGAGTACACCTT 1277

QY 1224 tcagaatagggccatacaaatcgtccatggtccatcagcattcttcccgactgcacaaac 1283
DB 1278 TCAAAACCTGCCTTATGACCTACCTCCTTGTGCTGTGTCAGCATTTCTTCTGACTGCANAA 1337

QY 1284 tggcgtttacaacactgcacaggttaactctcaaaagctcgaagcataaagatgcagctgc 1343
DB 1338 TGTAGTTTCAACACTGCANAAAGTGTGTTTCAAAAGGCTCGCTAGCAAAAGATGATTTGCTGT 1397

QY 1344 aggtgggtgattctcttgccagtcatacaatgaagaaacgcctactcgtgtagacagga 1403
DB 1398 CAACAGTGCATTTCTTGGCAGTCTGACACGAAGAAACACCTTCCGCAAAATTTATGATGC 1457

QY 1404 tacacttacagctaacgactatgggaacagaaaaacgtcaagaagattcatcagacta 1463
DB 1458 TGTATTTACCAAAAGATGGCTGTGGGAACAGATAAAGTGTACCAGAGATGCTACAGATTA 1517

QY 1464 tctgtgtacatgacaaatgtaaatatagatctaaagaaggatttctaagaagacggaaa 1523
DB 1518 CTTGTGTTATATACAGATGTGACAAATAGGTCCTCATGAAGCAATCTTTGAAGAATGGGCA 1577

QY 1524 ggaacttatcactcgtgtatgtccgctggtcgtctcgtctcgtcttctcgtcaatgaaa 1583
DB 1578 AGATCCCAATTTGACAGTCATGTGAGCGGCCATGCTTTTGTCATGTTTTTGTGAATGTGCA 1637

QY 1584 actatcagggaactgtttatgggtacattggataatccaaaacttacatcagtggaacgt 1643
DB 1638 ACTATCAGGAACCTGATATGACCAATTTGGAATAATCCCAAACTAGCTTTTAGTGGCAAGT 1697

QY 1644 gaagttaagagctggattataacaagatttctcgtcagtggttccgtgtggtcctccogaa 1703
DB 1698 GAAACTGAGAGCAGGAGTCAACAAGGTTTCTTTACTAAGTATCGCTGTGGCTTCCGAA 1757

QY 1704 cgttggcgtgcatatgatatacatggaatgcaagaggttcttaggtccagtcagcttgagcgg 1763
DB 1758 TGTTGGCTTACACTTTGAAACATGGAATGCTGGGGTCTTGGGTCCAGTGACATTTGAAAGG 1817

```

```
Qy 1764 tctcaatgaagggtcaagaaacttggcgaacacagaaaatgggtcttcaaggggtgtgctcga 1823
Db 1818 GGTGATTCAGCAACATGGGATATGCTAAAATGGAAATGGCTTTTACAAGATGGTCTGAA 1877
Qy 1824 agcgaatcgttaagctcttcaactccttaagtggaagttctctgttgtaatgggttcgagg 1883
Db 1878 AGCGGAACGCTTGAGCCCTTCATACAGTTAGTGGCAGTTCGCTGTGTGAGTGGGTGAAGG 1937
Qy 1884 ttcaataatggctcaaaagcagccctgacttggtcacaaggctacatttaacgqccctgg 1943
Db 1938 ATCATTAAGTCTCAAGACACACCCTCATTTGGTACAAAGACTACTTTTACGACCACT 1997
Qy 1944 aggaatgatccactagcttttagacatggcgaagtatgggaagaagtcagatatgataaa 2003
Db 1998 AGGTAAATGATCCATTAAGCTTTAGATATACAGATATGGAAGGATGGAATATGATGATAA 2057
Qy 2004 tggtagaagcgttaggtccatggcctggatatacatagacacaaagcgactcgacaaaatg 2063
Db 2058 TGGTCAAAAGTATTGGACGCCACTGGCCCTGGATATAAAGCTCGTGAAGTTGTGTGCTTG 2117
Qy 2064 cagttatgctggaacgttcaacgagaagaagtgcggaactaactcgcgacacactcttca 2123
Db 2118 CAACATGCTGGAATATATGATGAGAAAATGTCATAGTAACGTGGAAAGGCTTCTCA 2177
Qy 2124 gagatggaccatgttccacgatcgtggtcgaacaaagtggaacttggtagtagtatt 2183
Db 2178 GAGATGGTACCATTGTTCTCGCTCGTGCCTCAACCCAACTCGCAACCTATTAGTTGTTTT 2237
Qy 2184 cgaagaatggaggaagtaaccacaggaattctctagtcaggaga 2229
Db 2238 TGAGAAATGGGGTGTGTATCCAAACAAAGATTCTTTGGTGAAAGA 2283

RESULT 5
AB046543 Pyrus pyrifolia pJPBGAL mRNA for beta-D-galactosidase, complete
LOCUS AB046543 2643 bp mRNA linear PLN 27-JAN-2001
DEFINITION Pyrus pyrifolia pJPBGAL mRNA for beta-D-galactosidase, complete
cgs.
ACCESSION AB046543
VERSION AB046543.1 GI:12583686
KEYWORDS
SOURCE Pyrus pyrifolia (cultivar:Housul) ripening fruit cDNA to mRNA,
clone:pJPBGAL.
ORGANISM Pyrus pyrifolia
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids I; Rosales; Rosaceae; Maloideae; Pyrus.
REFERENCE 1 (sites)
Tateishi,A., Inoue,H., Shiba,H. and Yamaki,S.
Molecular cloning of beta-galactosidase from Japanese pear (Pyrus
pyrifolia) and its gene expression with fruit ripening
Unpublished
2 (bases 1 to 2643)
Tateishi,A., Inoue,H., Shiba,H. and Yamaki,S.
Direct Submission
Submitted (25-JUL-2000) Akira Tateishi, College of Bioresource
Sciences Nihon University, Lab. of Pomology and Vegetable Crops
Science, 1866 Kameino, Fujisawa-city, Kanagawa 252-8510, Japan
(E-mail:tateishi@brs.nihon-u.ac.jp, Tel:81466843622,
Fax:81466843622)
FEATURES
            location/Qualifiers
source 1..2643
            /organism="Pyrus pyrifolia"
            /cultivar="Housul"
            /db_xref="taxon:3767"
            /clone="pJPBGAL"
            /tissue_type="fruit"
            /dev_stage="ripening"
            183..2378
            /gene="pJPBGAL"
            183..2378
            /gene="pJPBGAL"
            /codon_start=1
```

```
/product="beta-D-galactosidase"
/protein_id="BAB21492.1"
/db_xref="GI:12583687"
/translation="MGVGIOTMWSILLFSCIPSAASASVSDHKAILINGOKRILIS
GSJHYPRSTPEMPPDLJOKADGGLDVQTVFVWNGHEPSPGKYFFEDRYDLVFKIL
VQAGLEVNLRIGIPYVCAEWNFGFPVWLAKYVPGIAFTDNEPEKAAQKQTEKIVSM
MKAELFQOTGGPIILISOIENEFPGVEIAGPKAYTKWAAQAVGLDGTGVPWIMCK
QEADPDVIDTCNGFYCENFKPKNDYKPMWTEVWTFTEFGGATYTRPADVAFSV
ARFIOSSGSLVNYMYHGGTNFRTAGGPMATSYDYADPLEYGLLRPKWGLHRLD
HKAIFKCESALVSDPSVTKLGSNOEAHVFKSESDDCAAFIANDYDKYSVKVSGGGQY
DLPWISILPDKTEVYSTAKVGSQSSQVQMPFVHSGFPWQSFIEETTSDEDTT
LDGLYEINIRDTDYLYMTDITIGSDEAFUKNGKSPLLTIFSAHALNVFINGQL
SGTVYGLENPKLFSQNVNLRSGINKLALLISVGLPNVGTHEFTWAGLVFITLK
GLNSGTWDMGSKWTKYTKGLKEALGLHTVTGSSVSEWVEGSPSKMAKQPTWYKATFN
APPCDAPLALDMGSMKGOTWINGSVRHWPGYIARGSCDCSYACTYDDKKCRTHC
GEPQRMWHIPRSWLTPNGLLVVFEEWGGDPSRISILVERGTALDAKKP"
polya_signal 2599..2604 532 c 622 g 704 t
BASE COUNT 785 a 532 c 622 g 704 t
ORIGIN
Query Match 49.6%; Score 1265.8; DB 8; Length 2643;
Best Local Similarity 74.0%; Pred. No. 1.4e-234;
Matches 1619; Conservative 0; Mismatches 567; Indels 3; Gaps 1;
Qy 113 tttttctcagtgaaagctagtgttctctatgatgacagagctataatcaataatgga 172
Db 235 TTTTCTTCGAGCTTCGGCTTCTGTGAGTTACGACCACAAAGCTATAATAATTAATGGC 294
Qy 173 aaagaaaaattcttattcttggtcaattcattatccaagaagcactccacagatggc 232
Db 295 AGAAAGGATTTTAATTTCTGGCTCCATCTACTATCCCAAGGACACTCTCAGATGTGC 354
Qy 233 ctgatctatacaaaagcctaaagatgaggttagatttatgaaacttatgtttct 292
Db 355 CGGATTTAATTCAGAGGCCAAAGATGGAGCTTGGATGTTATACAGACCTATGTGTTT 414
Qy 293 ggaatggacatgagcctctcctcgtgaaataataattttgaaggagaatgatctgtta 352
Db 415 GGAATGGCCAGCAACCTTCTCCGGGAAATAATATTATTCGAGGACAGATATGATTGGTCA 474
Qy 353 gattcatacaaatggtacaaagacagagactttatgcaatttcagttatggccctacg 412
Db 475 AGTTTCATCAAGCTGGTGCACAAAGCAGCCCTATTGTTAATCTCCGGATGGCCCTTATG 534
Qy 413 tctgtcgaatgaactttgggattccctgttttgctaaataatgtgcttgatgg 472
Db 535 TTTCGCTGGAATGGAACATTCGGGGATTTCCAGTTTGGCTGAAATATGTCCTGGAATCG 594
Qy 473 aatttagaacaacaatacagccttttaaggctggtatgcaaggattgttcagaaaaatag 532
Db 595 CTTTTCGAACGCAATGAGCCCTTCAAGCGCGCAATGCAAAAATTTACAGAGAAGATTG 654
Qy 533 tcaacatgatgaagtcagaaaaattgtttgaaatcctcaaggaggagccaataattatggccc 592
Db 655 TCAGCATGATGAAGGCGAGAGAAGCTGTTTCAAACTCAAGAGAGCTCTATAATTTCTCTC 714
Qy 593 agatacaaaatgagtatggaccagtagaattggaaattggtcctcctggtaaacctata 652
Db 715 AGATAGAAAATGAATTTGGACCTGTGNAATGGGAAATTTGGTCCGCTGGAAAAGCTTACA 774
Qy 653 caaatgggcagctcaaatggctgtaggtttgaaaaactgggtgtcccatggatcattgta 712
Db 775 CCAATGGCAGCTCAGATGGCTGTAGCTCTAGACACTGGAGTTCCATGGATTATGTGCA 834
Qy 713 agcaagagatgctcctcgtatcctgtgattgatacttgaatggtcttactcgaaaggt 772
Db 835 AGCAAGAGGATGCCCCCGATCCCTTTATTGACACTTGAATGGTTTCTACTGTGAGAAAT 894
Qy 773 tccgtcctaataagccttacaaacctaaatgtggacagaagatggaactggtggtgata 832
Db 895 TCAAGCCAAATAGGACTATTAAGCCCAAAATGTGGACAGAACTCTGGACTGTGTTGTTATA 954
```


SGTVYVGSLENPKLFSFQNVNLRSGINKLALLISVGLPNVGTHTFETWAGVLGPITLK
GLNGTMDMSWKYTYTGLKGEALGHTVSSVSEWQPSMAKQPLTWYKATFN
APPDAPLALDMSGMKGQIWIINGSVGRHWPVGIARGSCDYSYAGTYDDKKRTHC
GEPQSRWHIPRSWLTPPTGNLLVVFEEWGGDPSSRISLVERGTALDAKKL"
244...2316

mat_peptide /evidence=experimental

3' UTR /product="unnamed"

polyA_site 2371..2616

BASE COUNT /note="putative"

ORIGIN 2616

763 a 521 c 623 g 709 t

Query Match 49.2%; Score 1256.2; DB 8; Length 2616;

Best Local Similarity 74.9%; Pred. No. 1e-232;

Matches 1586; Conservative 0; Mismatches 528; Indels 3; Gaps 1;

QY	113	ttttttcttcagtgaaagctagtttcttcttatgatgacagagctataatcataatggga	172
DB	227	TTTTTTCTGCGAGCTTCAGCTTCGTGAGTTATGACCAAGGCTATAATAATTAAATGGGC	286
QY	173	aaagaaaaattcttattcttggttccaattcattatccaaagaagcactccacagatggc	232
DB	287	AGAAAAGGATTTTAATTTCTGGCTCCATTCACATATCCAGAAAGCACTCCCGAGATGTGC	346
QY	233	ctgatcttatcaaaaggctaaagatgagggcttagattgattgaacttatgtttct	292
DB	347	CGGATTTAATTCAGAAAGCCAAAGATGAGGCTTGAGTTGTATACAGACCTATGTGTTT	406
QY	293	ggaatggacatgagccttctctctggaaataataatttgaaggaagatatgatcttgta	352
DB	407	GGAATGGCCATGAACCTTCTCGGGAATATTATTTCGAGAAAGATATGATTTGGTCA	466
QY	353	gattcataaaatggtacaaagacagagcactttatgcaaatgaattatggcccttacg	412
DB	467	AGTTTATCAAGCTGTGTCAACAAGAGCCCTATTTTGTAAATCTCGGATTTGCCCTTATG	526
QY	413	ctgtgctgaatggaaacttggggatccctgttggctaaatatgtgctggtatgg	472
DB	527	TTTGGCTGAATGGAACTTCGGGGATTTCCAGATTTGGCTGAAATATGTCCCTGGAATCG	586
QY	473	aatttagaacaacaatcagcctttaaagggtgctcaaggattgttcagaaaaatag	532
DB	587	CTTTTCGAAGCGACAATGAGCTTCAAGGGCGCAATGCAAAAATTTACAGAGAAGATTG	646
QY	533	caacatgatgaatcagaaaaattgttgaatcattcaagggaggaaccaataattatggcc	592
DB	647	TCAGCATGATGAAGCGCAGAGAAGCTGTTTCAAACTCAAGGAGGTCTATAATTCTCTCTC	706
QY	593	agatacaaaatgagtagtaccagtagaatgggaattggtctcctggttaaaagcttata	652
DB	707	AGATAGAAAATGAATTTGGACCTGTGGAAATGGGAATTTGGTGACCTGGAAAAGCTTACA	766
QY	653	caaaatgggcagctcaaatggctgtagtttgaaaactgggtgtcccatggatcatgtgta	712
DB	767	CCAAAATGGCGAGCTCAGATGGCTGTAGGCTAGACACTGGAGTTCCATGGATTATGTCA	826
QY	713	agcaagagatgctcctgatcctgattgattgatacttgaattgaattcctactcgaagggt	772
DB	827	AGCAAGAGGATGCCCGGATCCGTTATGACATTTGCAATTTGCAATTTGCTTACTGTGAAAT	886
QY	773	tcctgcctcaataagccttacaaacctaaatgtggacagaagtatggactggctgtgata	832
DB	887	TCAAGCCAAATAAGGACTATAGCCCCAANAATGTGGACAGAGCTCTGGACTGGTGTGATA	946
QY	833	cgaaatctgggtgtccaatctctcaaaaccagccgaagacattgcatttccagttgcca	892
DB	947	CAGAATTCGGTGGGCGAGTTCCCACTAGACCTGCAAGAGATGTGCAATTTTCAGTTGCTA	1006
QY	893	ggttggctcagaacaatggttcattcttcaattactatcatatgatggaggaacaatt	952
DB	1007	GGTTTCATAAAGCGGGCTTCGTTTTTGAACATATTACATGTATACCAGGAGGAACGAAT	1066

QY	953	ttggccggacatcatcaggggttttcttcattgcaactagctacgatttatgatgctcctctg	1012
DB	1067	TTGGCCGAACAGCCGGAGGTCCCTTCATGGCCACTAGCTATGACTACGACGCCCTTAG	1126
QY	1013	atgaatatgggttgctgtaataaacaagtatgggcacttgagagacttacataaagcta	1072
DB	1127	ACGAATATGGACTACCCCGGAACCAAAAGTGGGACATTTGAGAGATCTGCACAAAGCA	1186
QY	1073	tcaagctatcgaacggcgttttagtttccatcatatgctgcgtgactagtctggaagta	1132
DB	1187	TTAAATCATGTAGTCTGCTTTAGTGTCCGTGTGACTTCTCAGTGACTAAACTCGGAAGTA	1246
QY	1133	atcaagagctcattgtttatagatcaaaatctggagctgtgctgtcttttttatccaaat	1192
DB	1247	ATCAGAGGCTCATGTATTCAANTCAGAGTC--GGATTGGCTGCATTCCTCGCAATTT	1303
QY	1193	atgactctagatattcagtaaaagtcacctttccagaataggccatacaaatctgctccat	1252
DB	1304	ATGACGCAAAATACTCTCTTAAAGTGAGCTTTGGAGGGCGGAGTATGACCTGCCGCAT	1363
QY	1253	ggtccatcagcattcttcccgactgcaaaactgcccgttttacaacactcacaggttaact	1312
DB	1364	GTCCATCAGCATTTCTCCGAGCTGCAAAACCGAAGTTTACAACACTGCAAAAGTTGGTT	1423
QY	1313	ctcaagctcagagcataaagatgacgctgctgcagggtgggtgattctcttggcagtcataca	1372
DB	1424	CGCAAGCTCGCAAGTTCAGATGACACACAGTACATAGTGGATTTCTTGGCAGTCATTCA	1483
QY	1373	atgaagaaagcctactctgatgacagcagatacaacttacagctaacggactatgggaac	1432
DB	1484	TCGAAGAAGACACGCTTCTGTATGAGACCGGATACAACTACATTTGACGGGTTGTATGAGC	1543
QY	1433	agaaaacgctcacaagagattcatcagactatctgtgtacatgacaatgtaaaatag	1492
DB	1544	AAATAAATATCATTACATAGGATATCAGACAGTACTTGTGTATGATGACATATCACAAATAG	1603
QY	1493	catctaaatgaagatttctaaagaacggaaaggaaccttctctcactgttatgtccgctg	1552
DB	1604	GTCTCTGTGAGAGCGTTCTTAAAGAAGCGAAAGTCCCCACTTCTTACGATCTTTTCAGCAG	1663
QY	1553	gtcagtcttgatctgtttctgcaatggaaactatcaggaactcttcttatggatcatg	1612
DB	1664	GTCTATGCTTGAATCTTATTCATCAATGTGTCAATGCTAGCACTAGGAACCGTGTATGGGT	1723
QY	1613	ataatccaaaacttacatcacagtgcaagtgcaagtgaaagttaagagctgttataacaagatt	1672
DB	1724	AGATTCCTTAATTTATCATTTCAAGTCAAAACGTTGAACCTGAGATCTGGCATCAACAACCTTG	1783
QY	1673	ctctgctcagtggttcccggttggtctcccgaaacgttggcgctgctattatgatacatggaatg	1732
DB	1784	CATTGCTTTAGCATTTCCGCTTGGTCTGCGCAATTTGTTGGTACTCACATTTTGAGACATGGAACG	1843
QY	1733	caggagttctcagggtccagtcacgttgagcgtctcctcaatgaagggttcaagaacttggcga	1792
DB	1844	CGGGAGTTCTTAGGCCCGATCATCGTTGAAGGCTCTGAATTCAGGAACATGGGACATGTGAG	1903
QY	1793	aacagaaatgctcttacaaggttgctgaaagcgcaatcgtaaagcttccactccttaa	1852
DB	1904	GGTGAATAATGACGTACAAAGACTGGTCTGAAAGGTGAAGCTTTAGGCCCTCCATCTGTTA	1963
QY	1853	gtggaggtcttctgttgaatgggttccaggttccactaactggctcaaaagcagcccttga	1912
DB	1964	CTGGGAGTTCTTCTTCAATGGGTAGAAGGAGCCATCGATGCTGAAAAACAACCCCTTA	2023
QY	1913	cttgtaacaaggctacatttaacggcgttgaggaataatgataccactagcttttagacaagg	1972
DB	2024	CATGTTACAAGGCTACTTTTAAATGCACCACCAGGCTGATGCTCCCAATTAGCTTTAGATATGG	2083
QY	1973	caagtatgggaaaaagctcagatagataaatggtgaagcggtaggtgcgcttggcctg	2032
DB	2084	GAAGCATGGGAAGGTCAGATATGGATTAATGGACAGAGCGTGGGACGCCACTGGCCCG	2143


```

Qy 2033 gatacatagcaagagcgactgcagcaaatgctcagttatgctggagcgttcaacgagaaga 2092
Db 2144 GGTACATTGCACGGCGGAGCTGTGGCGATTGTTCTTATGTCGGAACTTATGATGATAAGA 2203
Qy 2093 agtcggaactaactgcggagcaaccttctcagagatggtaccatggttccacgatcgtggc 2152
Db 2204 AATGCAGAACCTCATTTGTGGCGAGCCCTCTCAGAGATGTCACACATTCTCTCGATCATGGT 2263
Qy 2153 tgaaccacagtggaactggttagtgatttcgaagaatgggaggttaattccacacagaa 2212
Db 2264 TGACCCCGACTGGGAATCTTTGGTGGTGTTCGAAGATGGGTGCTGATCCGTCAGGA 2323
Qy 2213 tttctctagtcaggaga 2229
Db 2324 TTTGCTGTTGTTGAAAGA 2340

RESULT 7
CAAJ5042 2783 bp mRNA linear PLN 30-MAY-2001
LOCUS Cicer arietinum mRNA for beta-galactosidase (clone CanBgal-3).
DEFINITION AJ005042
VERSION AJ005042.2 GI:14274980
KEYWORDS beta-galactosidase.
SOURCE chickpea.
ORGANISM Cicer arietinum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids 1; Fabales; Fabaceae; Papilionoideae; Cicereae;
Cicer.
REFERENCE 1 (bases 1 to 2783)
AUTHORS Esteban,R., Dopico,B. and Labrador,E.
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 2783)
AUTHORS Labrador,E.
TITLE Direct Submission
JOURNAL Submitted (01-APR-1998) Labrador E., Dpto. Fisiologia Vegetal,
Univ. Salamanca, Campus Miguel de Unamuno. Pza. Doctores de la
Reina s/n, 37007-Salamanca, SPAIN
REMARK 3 (bases 1 to 2783)
AUTHORS Dopico,B.
TITLE Direct Submission
JOURNAL Submitted (16-SEP-1998) Dopico B., Dpto. Fisiologia Vegetal, Univ.
Salamanca, Campus Miguel de Unamuno. Pza. Doctores de la Reina s/n,
37007-Salamanca, SPAIN
REMARK 4 (bases 1 to 2783)
AUTHORS Dopico,B.
TITLE Direct Submission
JOURNAL Submitted (29-MAY-2001) Dopico B., Dpto. Fisiologia Vegetal, Univ.
Salamanca, Campus Miguel de Unamuno. Pza. Doctores de la Reina s/n,
37007-Salamanca, SPAIN
COMMENT On May 31, 2001 this sequence version replaced gi:3641862.
FEATURES
source
1. .2783
/organism="Cicer arietinum"
/cultivar="Castellana"
/db_xref="taxon:3827"
/clone="CanBgal-3"
/tissue_type="etioloated epicotyls"
/clone_lib="CAN-5"
/dev_stage="5-day-old seedlings"
147. .2339
/EC_number="3.2.1.23"
/codon_start=1
/product="beta-galactosidase"
/protein_id="CAA06309.1"
/db_xref="GI:3641863"
/db_xref="SPRMBL:065736"
/translation="MESLIRWRKRYVMVIGLVFLICLFVFSVTSVYDHRKAIYINGOR
RILISGSIHYPRSTQPMWDLIQAKDGGVDVIQTYVFWNGHEPSPGNYYFEDRFDLY
KFVKVVOAGLYVNLRIQYVCAEWNFGGFPVLMKYVPGVAFRTDNEPFRKAAMQKFTA

```

```

KIYVMMKAENLFESQGGPIIIMSIENEYGPWEIEIGAPKAYTKWFSQMAIGLDTGYP
WIMCKDEADPPIIDTNGYECNFENPKNYKPMWTEENSGWYTDGSAVYRPAOD
VAFSVARFIQNRGSVYVNYVHGTFNFRSTAGLFIATSYDYDAPIDEYGLISPKWG
HLNRLHKAIKOCPIILVSDPTYSWPKNLELVHYVKTSTGACAFANLANYDTSPAKVT
FGNQYDLDPFWSISILPCKTAVNTAKVTPSFHRMTPVSAFQWQSYNEAPASS
GIDSTTANALLEQIKVTRDSDLYMTVDVNSPNEGFIKNGQIPVLYTAMSAGHVLH
GVFNQGSFGTAYGLENPKLTFNSVKLRVGNKIKLSLVAAGLSNGLYHETWNVGV
LGPVYTLKLNEGTRDLGSKQWSYKIGLKGTELNLHTLIGSSVOWTKGSLVKKQPIU
WYKATPDAGNDPLALDMSMGKEIWNAGESIGRWHPAYIARGSCGGCNYAGTFTD
KKCTSCQGPQTKWYHTIPRSWNPGRNLFVLLEBGGDPSCISLVKKT"
BASE COUNT 874 a 484 c 619 g 806 t
ORIGIN
Query Match 48.2%; Score 1231.2; DB 8; Length 2783;
Best Local Similarity 73.0%; Pred. No. 6.7e-228;
Matches 1597; Conservative 0; Mismatches 588; Indels 3; Gaps 1;
Qy 51 gaatgtggaaaaaatgctaaggactaatgtgtgtgtgtatttagttattgtttatgga 110
Db 152 GAGTTTAAACAAGGTGGAGAAAGTTTATATGTTATAGGATTGGGTGTTTGTGTTT 211
Qy 111 tttttttcttcgaatgaagctagtggttctctatgatgacagagctataataaatcgg 170
Db 212 GTTTGTTTTTCTGTGTACAGCTTCTGTGACTTATGATCATAAAGCTATTGTAATATG 271
Qy 171 gaaagaaaaattcttctgttcaattcaattatccaaagaagcactccacagatg 230
Db 272 ACAGAGAAGAAATTTGATCTCTGGCTCTATTCTATTCATTCACGAAGCACACCTCAGATG 331
Qy 231 gcttgatcttatcaaaaaggctaaagatggaggttagatgttatgaaacttatgtttt 290
Db 332 GCCAGATTTAATTCAAAAGGCCAAGATGGAGGTGTTGATGTTATTCAGACATATGCTT 391
Qy 291 ctggaatggacatagcgtctctcctcgaaaaataataattttgaagagaagatgatcctcgt 350
Db 392 TTGGAATGCATGACATGAGCCTTCTCTCGAAATATTACTTTGAAGATAGGTTTACTTGT 451
Qy 351 tagattctcaaaatggtacaaagagcagactttatctcaattacattatggccctta 410
Db 452 TAAGTTTGTCAAGGTGGTTCACACAGCTGGCCTTTATGTTTAACTTGCCTATTGGTCCCTA 511
Qy 411 cgtctgtgctgaattggaaactttgggggattccctgtttgtggctaaataatgtgcctgggtat 470
Db 512 CGTTTGCCTCAATGGAACCTTTGGGGGATTTCTCTTTGGCTAAAGATATGACCAGGTGT 571
Qy 471 ggaatttagaacaacaatacagccttttaaggtggcgtgcaaggattgttccgaaaaat 530
Db 572 TGCCTTTCAGAACAGACAAACGAGCGCTTCAGGCGACCAATGCAAAAATTCACCTGCTAAGAT 631
Qy 531 agtcaaatgatgaatgcagaaaaattgtttgaaatctcaaggagagcaataaattatggc 590
Db 632 TGTACAGATGATGAAGCAGAGAAATTTATTTGAAGCCCAAGGTGCTCCAAATTAATATGTC 691
Qy 591 ccagatacaaaatgagtatggaccagtagaattgggaaaattggtgctcctggttaagctta 650
Db 692 TCAGATTGAGAATGAGTATGACCTGTGGAATGGGAAATTTGCTCTCTCGAAAAAGCTTA 751
Qy 651 taaaaaatggcagctcaaatcgtgtgtaggtttgaaaaactggttcccatggatcgtg 710
Db 752 CACCAAAATGGTTTCTCAATATGGCCATTGGTCTAGATACCGCGCGTACCGGTGATTATGTG 811
Qy 711 taagcaagaggtgctcctcgtatcctgtgattgatacttgtaattggtccttactcgaagg 770
Db 812 TAAGCAAGAAGACCGCTCTCTGATCTATTATTGATACCTTGCATGGATGATATTACTGTGAAA 871
Qy 771 gttcgtctctaataagccttacaacctaaatgtggcacagaagtatgacctggtcgtgta 830
Db 872 CTTTACTCCCCAACAGAACTACAAACCCAAAATGTGGGACTGAAAATTTGGCTGCTGCTGCTA 931
Qy 831 tacgaaattcgggtggtccaatctcctcaagaccagcgcaagacattcatttcagttgc 890
Db 932 CACTGATTTTGGGAGCGCGCTCCCTTATAGACCGGCACAGACGCTGGCATTCTCAGTTGC 991

```



```
QY 891 caggttctcagacaacatggttcattcttcttaattactacatgatatcatggaagacaacaa 950
    || || || || || || || || || || || || || || || || || || || || || || ||
Db 992 AAGATTACAAAATCGCGGCTCATATGTTAACTACTATATGCTACCATGGTGGAACTAA 1051
    || || || || || || || || || || || || || || || || || || || || || || ||
QY 951 ttttgcggacacatacagcgtttttcatttgcaactagctacgattatgatgctcctct 1010
    || || || || || || || || || || || || || || || || || || || || || || ||
Db 1052 TTTTGAAGGACATCGCTGGCCCTTTTCATTCGCCAAGCTATGACTAGACGTACATAT 1111
    || || || || || || || || || || || || || || || || || || || || || || ||
QY 1011 cgaatgaatggttctgtaagaaacaaagtatggcgacttgagagacttacataaagc 1070
    || || || || || || || || || || || || || || || || || || || || || || ||
Db 1112 TGATGAATACGGACTTCTAAGTGAACCAAAATGGGGCATCTAAGAAATTTACATAAAGC 1171
    || || || || || || || || || || || || || || || || || || || || || || ||
QY 1071 tatcaagctatctgaaccggcttttagtttctcatcatatctgctggtgactagtcttggag 1130
    || || || || || || || || || || || || || || || || || || || || || || ||
Db 1172 AATAAAGCAATGTGAGCCGATTTTAGTGTCTGTGGATCGGCGGTGTCATGGCCAGGAAA 1231
    || || || || || || || || || || || || || || || || || || || || || || ||
QY 1131 taatcaagaggtctatgtttatagatacaaaatctggagcttgtgctgttttttatceaa 1190
    || || || || || || || || || || || || || || || || || || || || || || ||
Db 1232 AAACCTTGAGGTACATGTTTACAAGACGACACTGGTGTGCTCGCTGTCATTCCTTGCAAA 1291
    || || || || || || || || || || || || || || || || || || || || || || ||
QY 1191 ctatgactatagatatcagtaaaagtcaccttcagaaatagggccatacaaatctgcctcc 1250
    || || || || || || || || || || || || || || || || || || || || || || ||
Db 1292 TTATGACACCATCTCCCGCAAAAGTTACATTTGGAATGGCCATAGCACCTACCGCC 1351
    || || || || || || || || || || || || || || || || || || || || || || ||
QY 1251 atggtccatcagcattcttcccgactgcaaaaactgcggtttacaacactgcacaggt--- 1307
    || || || || || || || || || || || || || || || || || || || || || || ||
Db 1352 TTGGGTCTATCAGTATTTCTCTGACTGCAAAACTGCTGTTTCAACACTGCAAAAGTTGG 1411
    || || || || || || || || || || || || || || || || || || || || || || ||
QY 1308 taactctaaagctcagacataaaagatgacgcctcaggtggtgattctcttgccagtc 1367
    || || || || || || || || || || || || || || || || || || || || || || ||
Db 1412 TAGAGTACCAAGCTCCATAGAAAAGATGACTCTCTGTAGTAGCGGATCGATTTGGCAGTC 1471
    || || || || || || || || || || || || || || || || || || || || || || ||
QY 1368 atacaatgaagaaacgctactgctgatgacagcgatacaacttacagctaacgcgactatg 1427
    || || || || || || || || || || || || || || || || || || || || || || ||
Db 1472 ATACAACGAAGCACCCGCTCTCGGTATAGATGATTTCCACCACGACGAAATGCATTTT 1531
    || || || || || || || || || || || || || || || || || || || || || || ||
QY 1428 ggaacagaaaacgtccaaagagatcatcagactatctgtgtggtacatgacaatgtaaa 1487
    || || || || || || || || || || || || || || || || || || || || || || ||
Db 1532 GGAGCAGATCAAGGTGACTCGCGATTTCTTCAGATTATTTTGGTATATGACAGATGTCAA 1591
    || || || || || || || || || || || || || || || || || || || || || || ||
QY 1488 tatagcatcctaagagatttctaaagacgaaagagatccttctcactgttatgtc 1547
    || || || || || || || || || || || || || || || || || || || || || || ||
Db 1592 CATTAGCTTAATGAAGGTTTATAAAGAAATGGTCAGTATATCTGTTCTTACTGCAATGTC 1651
    || || || || || || || || || || || || || || || || || || || || || || ||
QY 1548 cgctggtcatgtcttgcatgtttctogtcaatggaaactcatcaggaaactgtttatggtac 1607
    || || || || || || || || || || || || || || || || || || || || || || ||
Db 1652 AGCTGGCCATGTTCTGCATGTTTTTGTAAATGCTCAATTTTCGGGAAGCTGCTTATGGAGG 1711
    || || || || || || || || || || || || || || || || || || || || || || ||
QY 1608 attgataatccaaacttacatacagtggaacgtggaagttaagtgagctggtattaaacaa 1667
    || || || || || || || || || || || || || || || || || || || || || || ||
Db 1712 ATTGAGAATCCTAAATTAACATTTAGTAATAAGTGTAAAGCTGAGGGTTGGAATAACAA 1771
    || || || || || || || || || || || || || || || || || || || || || || ||
QY 1668 gatttctctgctcaggtttccgttgggtcccggaacgttggcgtggaattatgatacatg 1727
    || || || || || || || || || || || || || || || || || || || || || || ||
Db 1772 GATTTCTTTACTTAGTGTGGCGTGGTCTCTCGAATGTGGCTTGCACATGTAACAAATG 1831
    || || || || || || || || || || || || || || || || || || || || || || ||
QY 1728 gaatgcaggagtctctaggtccagtcacgttgagcggtctcgaatgaaggtcaagaacct 1787
    || || || || || || || || || || || || || || || || || || || || || || ||
Db 1832 GAATCTTGGTGTCTTAGGCCCGTAACGTTGAAGGCTCTAAACGAGGGGACAAAGAGACTT 1891
    || || || || || || || || || || || || || || || || || || || || || || ||
QY 1788 ggcgaacagaaaatggtttacaagagtttggttgcgaagcgcaatcggttaagtcctcaetc 1847
    || || || || || || || || || || || || || || || || || || || || || || ||
Db 1892 GTCTGGACAGAAATGGTCTTACAAGATTGGACTGAAGGGGAAACCTTGAACCTTTCATAC 1951
    || || || || || || || || || || || || || || || || || || || || || || ||
QY 1848 cttaaagtggaggttctctgttgtaaggttgcgaagttcactaaatggtcctcaaaacagcc 1907
    || || || || || || || || || || || || || || || || || || || || || || ||
Db 1952 CTTAATTGGAAGTCTCTCTGTTCAATGGACAAAAGGATCATCGTTGGTGAAGAAAACACC 2011
    || || || || || || || || || || || || || || || || || || || || || || ||
QY 1908 cctgacttggtacaaaggtctacatttaaagcgctgagagaaatgatccactagctttaga 1967
    || || || || || || || || || || || || || || || || || || || || || || ||
Db 2012 TTTGACTTTGGTACAAGCAACTTTTCGACGCCCCCGCTGGCACGATCCATTGGCGCTAGA 2071
    || || || || || || || || || || || || || || || || || || || || || || ||
```

```
QY 1968 catggcaagtatgggaaaaaggtcagatatggtataaattgdtgaagcgtaggtcgccattg 2027
    || || || || || || || || || || || || || || || || || || || || || || ||
Db 2072 TATGAGTAGCAGGGGAAAAAGGTGAAATATGGGTAAATGCTGAAAGCATTTGGTCGGCATTG 2131
    || || || || || || || || || || || || || || || || || || || || || || ||
QY 2028 gcttgatatacagacaagcgactgcaagaaatgcaaatgcaattatgctggaacgttcaacga 2087
    || || || || || || || || || || || || || || || || || || || || || || ||
Db 2132 GCCGCATATATAGCACGTGGGAAGTTGTGGTGTGTAATTATGCGGGAACGTTTACCAGA 2191
    || || || || || || || || || || || || || || || || || || || || || || ||
QY 2088 gaagaagtgcggaactaaactcggacaacctctctcagagatggtaccattgttccacgac 2147
    || || || || || || || || || || || || || || || || || || || || || || ||
Db 2192 TAAAAAATGCCCAACAAGCTGTGGACAACCCACCAAAATGGTATCATATTTCTCGCTC 2251
    || || || || || || || || || || || || || || || || || || || || || || ||
QY 2148 gtyggtgaaaccaaagtggaacctgttagtagtattctcgaagaatggggaggttaatccaac 2207
    || || || || || || || || || || || || || || || || || || || || || || ||
Db 2252 CTGGGTGAACCCCAAGAGGGAACCTTCTTAGTTGCTTGAAGAATGGGAGCGGATCCTTC 2311
    || || || || || || || || || || || || || || || || || || || || || || ||
QY 2208 aggaatttctctagtcaggagatcaaga 2235
    || || || || || || || || || || || || || || || || || || || || || || ||
Db 2312 TCGAATTTTCATTGGTGAAGAAAGAACATGA 2339
    || || || || || || || || || || || || || || || || || || || || || || ||
```

```
RESULT 8
CAR011010
DEFINITION
Cicer arietinum mRNA for beta-galactosidase, clone CanBgal-4.
ACCESSION
AJ011010
VERSION
AJ011010.1 GI:3641864
KEYWORDS
beta-galactosidase.
SOURCE
chickpea.
ORGANISM
Cicer arietinum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Cicereae;
Cicer.
REFERENCE
1 (bases 1 to 2821)
Labrador, E.
Direct Submission
Submitted (14-SEP-1998) Labrador E., Dpto. Fisiologia Vegetal,
Univ. Salamanca, Campus Miguel de Unamuno. Pza. Doctores de la
Reina s/n, 37007-Salamanca, SPAIN
2 (bases 1 to 2821)
Dopico, B., Esteban, R. and Labrador, E.
The fourth beta-galactosidase clone found in Cicer epicotyls
Unpublished
FEATURES
Location/Qualifiers
1..2821
/organism="Cicer arietinum"
/cultivar="castellana"
/db_xref="taxon:3827"
/clone="CanBgal-4"
/tissue_type="etiolated epicotyls"
/clone_lib="CAN-5"
/dev_stage="5 days old seedling"
204..2375
/EC_number="3.2.1.23"
/codon_start=1
/product="beta-galactosidase"
/protein_id="CAA09457.1"
/db_xref="GI:3641865"
/db_xref="SPTREMBL:O82670"
/translaton="MRKEFNLLGLMLLCFVWCVAVTASVYDHTKIIVIDGQRLLISG
SIHYPRSTPEMPALFQKAKEGGLDVQTYVFNWHEPSPGKYTFEDFLDKFKILA
QOAGLYVHLRIGPVCAEWNFGFPVWLKVPVPGISFTDNEPFAAKQKFTTKIVSMX
KAENLFNOGPIIMSQIENEYGPVENIGNAPGKAYTNAAQMAVGLDTGVPDMCKO
EDAPDPVIDTCNGYICNFENPKNKPKMWTENWSGMYTDFGNAICYRPVEDLAYSVA
RFTONRGSFVNYMYHGGTNGFTSSGLFIATSYDYDAPIDEYGLTNEPKWSHRDLH
RAIKQCEPALVSVDPTITSLGNKLEAHVYSTGVSACAAFLANDYTRKSAATVFNFGKY
DLPPWSVSLPDKCTDVFNTAKVGAQSSQKTMISTNSTFDWQYIEEPATSSSEDSIT
AELWAEQINVTROSSDYLVLTIDVNIISPNEFD IKNGQYPIILNVMSAGHVLHVPVNGQL
SGTYGVLDNPKLTFNSVNLTVGNKKISLLSVAVGLPNVGLHPETWNVGVGLPVTIKL
GLNGETRDLSNQKWSYKVKGLKGESLSLHTITGGSSVDWDTGSLAKKQPLTWYKATN
APAGNDPLGLDMSMGKEIWNDOSTIGRHPWGYIAHGSGDCDGYAGTFTNTKCRNTC
GNPTQTWTHIPRSLNLTGNVLVVEEWGGDPSGISLLKRI"
```

CDS

Db	2203	GTGATTATGCTGGGAACATTTTACTAATACAAAATGTAGACAAAATGTGGAAATCCCAAC	2262
Qy	2123	agagatggtaccatgttccacgacgtcggtcggaacaaagtggaactttagtagtat	2182
Db	2263	AAACATGGTATCATATTCCTCGCTCATCGGTGTGATCAACTGGGACGCTCTGGTTGTGT	2322
Qy	2183	tcgaagaatggggagggttaataccaacagagaattctctagtcagagatcaagat	2236
Db	2323	TGGAAGAATGGGAGGTGACCCCTCTGGAATTTCTGCTCAAAAGAATATGAT	2376
RESULT	9		
AF229794			
LOCUS	AF229794	2166 bp	mRNA linear PLN 03-MAY-2000
DEFINITION	Vigna radiata beta galactosidase mRNA, complete cds.		
ACCESSION	AF229794		
VERSION	AF229794.1	GI:7682676	
KEYWORDS			
SOURCE	Vigna radiata.		
ORGANISM	Vigna radiata		
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Vigna.		
REFERENCE	1. (bases 1 to 2166)		
AUTHORS	Chen, M.-C., Li, S.-C., Chen, K.-C. and Chen, C.-S.		
TITLE	Direct Submission		
JOURNAL	Submitted (30-JAN-2000) Institute of Botany, Sinica, Taipei, Taiwan 11529, Republic of China		
FEATURES	Location/Qualifiers		
source	1..2166		
	/organism="Vigna radiata"		
	/db_xref="taxon:157791"		
CDS	1..2166		
	/codon_start=1		
	/product="beta galactosidase"		
	/protein_id="AAF67341.1"		
	/db_xref="GI:7682677"		
	/translation="MGKGEHGGVLMLLFFWCGVTVASVYDTHKAIVIDGKRRILLISL SHYRSTPQMPDLLOKAGHDGLVQITVFNWGHEPSPGYKFFIEDRLDLVRFVLA QAGLVYHLRIGYICAEWNFGYVWLYKVPATFATNEPFAAMQKFTAKIVISM KEERLFSQGGPILISQIENYGFVWEIGAPGKSYTKWAAQMAVGLDTGVPWMVKQ RFAQDPVIDTCNGFYCNFKPNKTKPKWNTENWTGYTDFGASPIRPAEDLAFVA RFTQDGSFVNYVMYHGGTNFGRTSGGLFTATSDYDADPLDEYGLONEPKWGHURLH KALQSEPALVTDPRKTVSLGVLEAHVFTPGCAAFATANYDTKSAKATPGSGUJL LPWQSIILPDCATVYNTARVGNWKKMTPVNSGFAQSYNEPASPSSQDSISARE ALMEQVITPQSDSILWYMTDVTYINGNEGLKNGRSPVLTYSAGHLHLVIFINQUSLQ TVYTGDRGNPKLTFSDNRVGNKSLLSAVGLPNVGVHFETWAGVLPVPTLKG L NCTTGLNQRKWSYKVLGKEALNLHTESSGSSEVITQGLSLVAKKQPLWYKATFSAP AGNDPLALDLGSMGKEGVNWSIGRHWPGYIAHSCNACNVAGYITTDQKCRITNGCK PSQRYVHVPWSLWNSGNSILVFEEMGDPNGIALVKRI"		
BASE COUNT	599 a	416 c	544 g
ORIGIN	607 t		
Query Match	47.3%	Score 1209;	DB 8; Length 2166;
Best Local Similarity	73.68;	Pred. No. 1.3e-223;	
Matches 1568;	Conservative	0; Mismatches 555;	Indels 6; Gaps 2;
Qy	103	ttatgtgatttttttcagtgaaagctagtgtttcttatgatgacagagctataatc	162
Db	40	TTGTTTTCTGGGTTTCTGGTGTCACAGCCTCTGCTACTTACGACCAACAAAGCATTTGTG	99
Qy	163	ataaatgggaaagaaaaattcttattcttggttcaattcaattatccaagaagcaactcca	222
Db	100	ATTGATGGAAGAGGAGAAATTTGATCTGGCTCCACTTTCACCTACCCAAGAAGCACCTCT	159
Qy	223	cagatggcctaatcttatacaaaagcctaaagatgagagcttagattgattgaact	282
Db	160	CAATGTGGCCAGACCCATTCAGAAGGCCAAGATGGAGCCTTGATGTCATTTCAGACC	219
Qy	283	tatgttttctggaatgacatgagcctctctcttggaataataatttttgaggagatat	342

QY	1423	ctatgggaaacagaaaaacgctcacaagagattccatcagactctctgtggtacatgacaaat	1482			
Db <td>1354</td> <td>CTCTGGGAACAGGTCATATGTACCCGGGATTTCTTCAGATTTATTTGTGTATATGACAGAT</td> <td>1413</td>	1354	CTCTGGGAACAGGTCATATGTACCCGGGATTTCTTCAGATTTATTTGTGTATATGACAGAT	1413			
QY <td>1483</td> <td>gtaaatatagcatctaatgaagatttctaagaacggaagagatcccttatctcaactgtt</td> <td>1542</td>	1483	gtaaatatagcatctaatgaagatttctaagaacggaagagatcccttatctcaactgtt	1542			
Db <td>1414</td> <td>GTCTACATTAAATGTAATGAAGGTTTTCTAAGAAATGTGTCGATCTCCTGTTCTTCACTGTA</td> <td>1473</td>	1414	GTCTACATTAAATGTAATGAAGGTTTTCTAAGAAATGTGTCGATCTCCTGTTCTTCACTGTA	1473			
QY <td>1543</td> <td>atgtccgctgctcatctgtctgtcatgtttctgcataatggaataactatcaggaaactgtttat</td> <td>1602</td>	1543	atgtccgctgctcatctgtctgtcatgtttctgcataatggaataactatcaggaaactgtttat	1602			
Db <td>1474</td> <td>ATGTCAGCAGGCCACTCTCTACATGTTTTCATTAATGGTCAACTTTCAGGAACGTGTAT</td> <td>1533</td>	1474	ATGTCAGCAGGCCACTCTCTACATGTTTTCATTAATGGTCAACTTTCAGGAACGTGTAT	1533			
QY <td>1603</td> <td>ggtcatttggaataatcaaaaacttacatcacagtggtggaacgtaagagctgggtatt</td> <td>1662</td>	1603	ggtcatttggaataatcaaaaacttacatcacagtggtggaacgtaagagctgggtatt	1662			
Db <td>1534</td> <td>GGGGGATGGGGAAATCCCTAAATTAACATTTAGTGACAAATGTCAACTTGAGAGTTGGCAAT</td> <td>1593</td>	1534	GGGGGATGGGGAAATCCCTAAATTAACATTTAGTGACAAATGTCAACTTGAGAGTTGGCAAT	1593			
QY <td>1663</td> <td>acaagatttctctgctcagtggtttccgttggtctccgaacggtggcgtccattatgat</td> <td>1722</td>	1663	acaagatttctctgctcagtggtttccgttggtctccgaacggtggcgtccattatgat	1722			
Db <td>1594</td> <td>AACAAGCTTTCTTTACTAAGTGTGTCCTGGTCTCCGGAATGTGGCGTGCATTTGAA</td> <td>1653</td>	1594	AACAAGCTTTCTTTACTAAGTGTGTCCTGGTCTCCGGAATGTGGCGTGCATTTGAA	1653			
QY <td>1723</td> <td>acatggaatcgagagttctaggtccagtcacgttgagcgtctcaatgaagggtcaaga</td> <td>1782</td>	1723	acatggaatcgagagttctaggtccagtcacgttgagcgtctcaatgaagggtcaaga	1782			
Db <td>1654</td> <td>ACATGGAATCGGGGTGTAGGCCAGTCTACTCTGAAGGCTCTAAATGAAGGAACACGA</td> <td>1713</td>	1654	ACATGGAATCGGGGTGTAGGCCAGTCTACTCTGAAGGCTCTAAATGAAGGAACACGA	1713			
QY <td>1783</td> <td>aactggcgaacagaaaatggtcttacaaggttggtctgaaaggcgaatcgttaagtctt</td> <td>1842</td>	1783	aactggcgaacagaaaatggtcttacaaggttggtctgaaaggcgaatcgttaagtctt	1842			
Db <td>1714</td> <td>GACTTGTCCGACAGAAATGCTCTACAAGGTAGGACTGAAAGGTGAAGCCTTGAACCTT</td> <td>1773</td>	1714	GACTTGTCCGACAGAAATGCTCTACAAGGTAGGACTGAAAGGTGAAGCCTTGAACCTT	1773			
QY <td>1843</td> <td>cactccttaagtggaggtctctgttgaaatgggttcgaggttcactaatggctcaaaaag</td> <td>1902</td>	1843	cactccttaagtggaggtctctgttgaaatgggttcgaggttcactaatggctcaaaaag	1902			
Db <td>1774</td> <td>CACACTGAAGTGGGAGTAGCTCTGTTGAATGGATACAAGGATCATTTAGTGGCTTAAAAA</td> <td>1833</td>	1774	CACACTGAAGTGGGAGTAGCTCTGTTGAATGGATACAAGGATCATTTAGTGGCTTAAAAA	1833			
QY <td>1903</td> <td>cagccctgacttggtaacaaagctacatttaacgcgcctggaggaatgattccactagct</td> <td>1962</td>	1903	cagccctgacttggtaacaaagctacatttaacgcgcctggaggaatgattccactagct	1962			
Db <td>1834</td> <td>CAACCTTTGACATGGTATAAGCGCAACTTTTAGTGCACCAGCTGGCAATGCCCATTTGCT</td> <td>1893</td>	1834	CAACCTTTGACATGGTATAAGCGCAACTTTTAGTGCACCAGCTGGCAATGCCCATTTGCT	1893			
QY <td>1963</td> <td>ttagacatggcaagtatgggaaaaggttcagatattgataaaatggtgaaaggcgtaggctgc</td> <td>2022</td>	1963	ttagacatggcaagtatgggaaaaggttcagatattgataaaatggtgaaaggcgtaggctgc	2022			
Db <td>1894</td> <td>CTAGATTTGGGTAGCATGGGAAGGGTGAAGTATATGGGTCAACGGTTCGACGATTTGGTGC</td> <td>1953</td>	1894	CTAGATTTGGGTAGCATGGGAAGGGTGAAGTATATGGGTCAACGGTTCGACGATTTGGTGC	1953			
QY <td>2023</td> <td>cattggcctgatacatagacaaggcagcctgcagcaaatgcagttatgctggaacgttc</td> <td>2082</td>	2023	cattggcctgatacatagacaaggcagcctgcagcaaatgcagttatgctggaacgttc	2082			
Db <td>1954</td> <td>CATTGGCCTGGATATATTGGACATGGGCAGTTGTATGCTTCGCAACTATGCTGGATATTAT</td> <td>2013</td>	1954	CATTGGCCTGGATATATTGGACATGGGCAGTTGTATGCTTCGCAACTATGCTGGATATTAT	2013			
QY <td>2083</td> <td>aacgagaagaagtcggaactaacctcggaacaccccttcagagatggtacattgttcca</td> <td>2142</td>	2083	aacgagaagaagtcggaactaacctcggaacaccccttcagagatggtacattgttcca	2142			
Db <td>2014</td> <td>ACCGATCAAAATCGCAGGACAAATTTGTTGGAACCAATCTCAGAGATGATATCATGTTCT</td> <td>2073</td>	2014	ACCGATCAAAATCGCAGGACAAATTTGTTGGAACCAATCTCAGAGATGATATCATGTTCT	2073			
QY <td>2143</td> <td>cgatctggttgaaaccaaactggtgaaactgttagtagtatttcgaagaatggggaggtaat</td> <td>2202</td>	2143	cgatctggttgaaaccaaactggtgaaactgttagtagtatttcgaagaatggggaggtaat	2202			
Db <td>2074</td> <td>CGTTCATGGCTGAATCGGGTGGTAACCTCTGTTGGTTCGTTTGAAGAATGGGAGGTGAT</td> <td>2133</td>	2074	CGTTCATGGCTGAATCGGGTGGTAACCTCTGTTGGTTCGTTTGAAGAATGGGAGGTGAT	2133			
QY <td>2203</td> <td>ccacagaagaattctctagtcaggagatc</td> <td>2231</td>	2203	ccacagaagaattctctagtcaggagatc	2231			
Db <td>2134</td> <td>CCTAATGGAATTGCTTTGGTAAAGAGAAC</td> <td>2162</td>	2134	CCTAATGGAATTGCTTTGGTAAAGAGAAC	2162			
RESULT 10						
LOCUS	LAN011047	2628 bp	linear			
DEFINITION	Lupinus angustifolius mRNA for exo galactanase.					
ACCESSION	AJ011047					
VERSION	AJ011047.1 GI:3860419					
KEYWORDS	exo galactanase.					
SOURCE	narrow-leaved blue lupine.					
ORGANISM	Lupinus angustifolius					
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;						
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;						
Rosidae; eustosids I; Fabales; Fabaceae; Papilionoideae; Genisteae;						
Lupinus.						
REFERENCE	1 (bases 1 to 2628)					
AUTHORS	Chengappa,S.					

```
Qy 518 ttgttcagaaatagtcacatgatgaagtcagaaatattgtttgaaatctcaaggagac 577
Db 614 TCACTGAGAGATTGTTAAATATAATGAAGACAGAGAAGTTGTTTCAATCCACGGAGGTC 673
Qy 578 caataaattatggccagatatacaaaatagatgtagacagtagaataaggaaattggtctc 637
Db 674 CAATAAATCTGCTCAGATAGAGAATGAGTAGACAGTGAATGGAAATGGTGCTC 733
Qy 638 ctggttaagcttatcaaaatggcgagctcaaatggctgtaggtttgaaactggtctc 697
Db 734 CTGGAAGAAGCTTATACCAATGGGCTGCTCAATGGCTGAGGTCTAGATCTAGTGTCC 793
Qy 698 catggatcatgttaagcaagagatgctcctgatctcgtgattgatacttgaatgct 757
Db 794 CATGGTATGTGCAAGCAAGAAGATCACCTGATCCTATTATTGATACCTGCAATGGAT 853
Qy 758 tctactcgaaaggttcgctcctaataagccttacaacctaataatgtggacagaatgat 817
Db 854 TTTACTGTGAAAACCTTCACTCCAACAAAGAACTACAACCCCAAAATTGTCGACAGAAAATT 913
Qy 818 ggaactggtgtatcacgaaattcgtggtccaattcctcaaaagaccagccgaagacattg 877
Db 914 GGACTGGCTGGTACATGCTTTTGGTGGTGAACCCCTTATGACCCAGACGAAGATATAG 973
Qy 878 cattttcaagttgcaggtttgttcagaaatggttcattcttcaattactacatgctatc 937
Db 974 CATTTTCAGTTGCCAGATTTCATTGAGAATCGCGGCTCACCTTTTAACACTATATGATC 1033
Qy 938 atggagagaacaaatttggccggagcatcatcaagggcttttcaattgcaactagctacatt 997
Db 1034 ATGGAGGAACATAACTTTGGCGGACATCGAATGGCCCTTCGTTGGCCACAAGTTATGACT 1093
Qy 998 atgaatcctctcgtatgaatatggttgcgtgaatgaacaaagatggccacttgagag 1057
Db 1094 ATGAATCTCCATTTGAATATGAGACTTCTAATGATGAACCAAAATGGGGGATCTGAGAG 1153
Qy 1058 acttacataaagctatcaagctatctgaacggcttttagtttcatcatatgctgcgtga 1117
Db 1154 AATTACATAGAGCAATAAAACAATGCGAGTGGGCTTTAGTGTGCTGGTACCCACAGTGT 1213
Qy 1118 ctgactcttggaagtaatacagaggtcctatttatagatacaaaatcgtggagcttgctg 1177
Db 1214 CATGGCTGGAAGAAACCTTTGAGGTAGATTGTTGTAACAAGACAGAGTCT---GCCTGTGCTG 1270
Qy 1178 cttttttatcaactatgactctagatctcagtaaaagtcacctttcagaataaggccat 1237
Db 1271 CATTTCTGCAAAATTATAACCCGACTATTCAACGCAAGTTAAATTCGGAAATGGACAAT 1330
Qy 1238 acaatcgtcctccatggttccatcagcattcttcccagctgcaaaactgcccgtttacaaca 1297
Db 1331 ATGATCTACCAACCTTGGTCTATCAGTATTCTTCTGACTGCAAAACCTGAAGTTTCAACA 1390
Qy 1298 ctgcacagggttaactctcaagctcgagcataaagatgacgctgcaagtggtgattct 1357
Db 1391 CTGCAAGAGGTTAATTCGCCAGATTACATAGGAAATGACTCCAGTAACAGTGCATTTG 1450
Qy 1358 cttgagcgtcatacaatgaagaacgcctactcgtgatgacagcgtacacacttacaacta 1417
Db 1451 CTTGGCAGTATACAATGAGAGACCTGCATCATCAACGGAATGATCCCGTCACAGAT 1510
Qy 1418 acggaactatgggaacagaaacgtcaagagattcattcagactatctgtggtacatga 1477
Db 1511 ATGCACATATGGAGAGGTTGGCGTGACCCGGATTTCTCCGATTATTGTTGGTACCTGA 1570
Qy 1478 caaatgtaatatagatcatcaatgaagatttctaaagaacggaagatcccttatctca 1537
Db 1571 CAGATGTCAACATTTGCTTAATGA-----TATAAGAGGATGGGAATGGCCTGTTCTGA 1624
Qy 1538 ctgttatgctcgtggttcattgctcttgatgttttcgtccaatggaaactacaggaactg 1597
Db 1625 CAGCAATGTCAGCAGGTGATCTTCTGATGTTTTCATCATGTTGTCATATATCCAGGAATG 1684
Qy 1598 ttatggtacattggataatcccaaaacttacatatacagtggaacogtgaaagtgctg 1657
```

```
Db 1685 CATATGGAGTCTAGATGATCCTAGATTAAACATTTAGTCAAAGTGTGAATCTGAGAGTTG 1744
Qy 1658 gtattacaagattctctcgtcgaagtgttcocgttggtctccccaagcttggtgctgcaatt 1717
Db 1745 GCAATAACAAGATTCTTTACTTAGTGTTCCTGTTGGTCTCGGAAATGTTGGTACTCACT 1804
Qy 1718 atgatacatgaaatgcaggagttcttagtccagtcacgctgagcggtctcctaataagaggt 1777
Db 1805 TTGAGACATGGAATCTGAGAGTCTTGGTCCAGTCACTACACTGACAGGTCTTAAGTAGCGGAA 1864
Qy 1778 caagaaacttgggaaacagaaatggtcttacaaggttggctgaaaggogaaatcgtaa 1837
Db 1865 CATGGGATCTTTCGAACAAAAATGGCTTACAAGATTGGTCTGAAAAGGTGAAAAGCTTGA 1924
Qy 1838 gtcttcaactctaagtggaagttctctgttgaatgggttcgaggttcactaatatggtc 1897
Db 1925 GCCPTTCACTGAAGCTGGGAGTAACCTGTTGAATGGGTACAAGGATCTTTAGTGGCTA 1984
Qy 1898 aaaagcagccctgacttggtaacaaggtctacaaaggtcatttaacgcgcctggaggaatgatccac 1957
Db 1985 AAAAAACAACCTTTGGCATGGTATAAGACAACCTTTTAGCGCACACCGCGCAACGATCCGT 2044
Qy 1958 tagctttagacatggcgaagtatgggaaaggtcagatatgataaaatggtgaaagcgctag 2017
Db 2045 TGGCTCTGGATTTAGTGTAGCATGGGTAAAGGTGAAGTATGGGTAAATGGTCAAAAGCATTG 2104
Qy 2018 gtgcattggcctggtatatacatagacaaagcgactcagcaaaatgcagttatctctgaa 2077
Db 2105 GACCCCATTTGCCCTGGAATAAAGCTCTGCTGTAATTCGGCAATTTGTAATTTACCTGGAA 2164
Qy 2078 cgttcaacgagaagaagtgcggaactaactgcgacacaccttctcagagatggtaccatg 2137
Db 2165 CTTATACCGATACAAATGCTTAGCAAACTGTGCAAAACCTCCCAAGATGGTATCATG 2224
Qy 2138 ttcaacgatcgtggtgaaacaaagtgtgaaacttggtagtagtattcagaagaatggggag 2197
Db 2225 TTCTCTCGTCTATGCTGAGATCGGTGCTAACTACTTGGTGTGCTAGAGAATGGGGAG 2284
Qy 2198 gtaatccaaacagaattctctagtcaggagatca 2232
Db 2285 GTGATCTAATGGAATTCGTTTGGTGGAAAGAAC 2319

RESULT 11
LOCUS A46218 Sequence 1 from Patent WO9523228. 2628 bp DNA linear PAT 07-MAR-1997
DEFINITION A46218
ACCESSION A46218
VERSION A46218.1 GI:2300464
KEYWORDS
SOURCE unidentified.
ORGANISM unclassified.
REFERENCE 1 (bases 1 to 2628)
AUTHORS Chengappa,S., Hellyer,S.A., De,S.J. and Reid,J.S.
TITLE NOVEL EXO-(1,4)- beta -D GALACTANASE
JOURNAL Patent: WO 9523228-A 1 31-AUG-1995;
UNILEVER PLC (GB)
COMMENT Other publication AU 1713795 950911.
FEATURES
source: Location/Qualifiers
/db_xref="taxon:32644"
/organism="unidentified"
/notes="unnamed protein product"
CDS
130..2322
/codon_start=1
/protein_id="CAA02881.1"
/db_xref="GI:2300465"
/translation="MFGSRVTMESLMRRNFHVLVLLFPWCYVTASVTDHKAIMI
NGORRLISGSIHPSTPQMPDLIQAKDGLDVIETVFWNGHEPSPKYYFEDR
FDLVGKIKLVQOAGLFLHURIGRIFCAENWFGFPVWLKVPGLAIFRTDNEPKFKAHQ
KFTEKIVNINKAEKLFQSGGPIILSIOIENEYGPVWEIGAPGPKATKWAQMAVGILD
```

Query Match 47.2%; Score 1205.8; DB 6; Length 2628;
Best Local Similarity 73.5%; Pred. No. 5.4e-223;
Matches 1569; Conservative 0; Mismatches 557; Indels 9; Gaps 2;
BASE COUNT 835 a 448 c 581 g 764 t
ORIGIN

1034 ATGGAGGAATAAATTTGGCGGACATGAATGGCCCTCTCTGTTGCCCAAGTTATGACT 1093
1098 atgatgctcctctcgaatgaatggttgcgtaaaagtaaaagtaaaagtaaaagtaaaag 1057
1094 ATGATGCTCCCATGATGAATATGGACTTCTAAATGAACCAAAATGGGGCATCTGAGAG 1153
1058 acttacataaagctcatcaagcttatctgaaccgctctttagtttccatcatatgctcggtg 1117
1154 AATACATAGACCAATAAACAATCCGAGTGGCTTTAGTGTGGTGGATCCACAGTGT 1213
1118 ctagtcttgggaagtaacagaggtcctatgtttatagatcaaaatctgagagcttgcgtg 1177
1214 CATGGCCTGGAAAAAACCCTTGAGGTACATTTGTCAAGACAGAGTCT--GCCTGTGCTG 1270
1178 cttttttatcaactatgactctagatattcagtaaaagtaaaagtaaaagtaaaagtaaa 1237
1271 CATTTCTTGGCAAAATTAACACCGACTATTCAACCAAGTTAAATTCGGAAATGGACAAT 1330
1238 acaatctcctccatggtcccatcagcattctcccgactgcaaaactccgcttttacaaca 1297
1331 ATGATCTACCACCTTGGTCTATCAGTATCTCTCTGACTGCAAAACTGAAGTTTCAACA 1390
1298 ctgcacaggttaactctcaagctcgagcataaaagatgacgctcgaggtggtgattct 1357
1391 CTGCAAAAGTTAAATCCCGAGATTACATAGGAAATGACATCCAGTAAACAGTGCATTG 1450
1358 cttggcagctacataaataagaacgctactgctgtagacagcagatacacattacageta 1417
1451 CTTGGCAGTCTACAAATGAAGAACCTGCATCATCAAGCGCAAAATGATCCCGTCACAGAT 1510
1418 acgactatggaacagaaaaacgtcacagagattcctcagactatcagactatctggttacatga 1477
1511 ATGACATATGAGGAGCAGGTTGGCGTGACCCCGGATCTCTCCGATTTATTTGGTGGTACCTGA 1570
1478 caaatgtaaatatagcatcctaatgaagatttctaaagacggaagatccttatctca 1537
1571 CAGATGTCAACATTTGGTCTTAATGA-----TATAAAGGATGGGAAATGGCGTGTCTGA 1624
1538 ctgttatctcgcgtggtcattgcttgcattgcttgcattgcttgcattgcttgcattgcttgcattg 1597
1625 CAGCAATGTCAAGCAGGTCATGTCTGAATGTTTTCATCAATGGTCAATATGTCAGGAACGTG 1684
1598 ttatggtacattgagataatcaaaactacacagtggaacgtggaacgtggaacgtggaacgtg 1657
1695 CATATGGAGCTCTAGATGATCCTAGATTTAATATTTAGTCAAAAGTGTGAATCTGAGAGTTG 1744
1658 gtattacaagatttctcgtcagtggttccgttggctcccgaaacgttggcgtgcatt 1717
1745 GCAATAACAAGATTCTTTACTTACTTGTTCGGTTGGTCTCGCGAATGTTGGTACTCACT 1804
1718 atgatacatggaatgcaggagttctaggtccagtcacgttgcagtcagtcagtcagtcagtcagtcag 1777
1805 TTGAGACATGGAATGACTTGGTGTGGTCCAGTCCACTGACAGGTCTAAGTAGCGGAA 1864
1778 caagaaactctggcgaacagaaaaatggtcttacaaggttggctgaaagcgaaatcggttaa 1837
1865 CATGGATCTTTCGAAGCAAAATGGTCTTACAGATTTGGTCTGAAAGGTGAAGACCTTGA 1924
1838 gtcttcaactccttaagtggtggttcttctgttgaatgggttcgaggttcacataatggctc 1897
1925 GCCTTCATACCTGAAGCTGGGAGTAACCTGTGTAATGGGTACAAGGATCTTTTAGTGGCTA 1984
1898 aaaagcagccctgacttgggtacaaagctacatttaacgcgccttggaagaaatgatccac 1957
1985 AAAAACAACCTTTGGCATGGTATAGACAACTTTTAGCGCACCGACCGGCAACGATCCGT 2044
1958 tagctttagacatggaagatggaagaaaggtcagatagataaaatggtaagcgctgag 2017
2045 TGGCTCTGGGATTTAGTACGATGGGTAAAGGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 2104
2018 gtcgccattggccttggatcatagcacaagggcagcagcaaatgagcttatgctgga 2077

QY 806 ggacagaagatgagtggtggtatatacagaataatcggtggtccaaatctcctcaaaagaccag 865
Db 976 GCACGAAATCTGGACTGCTGATACAGAAATTTGGTGGGCGAGTCTCTACTAGACCTG 1035
QY 866 ccgaagacattcatttccagttgcaggttgggttcgaagaacaatggttcaatcttcaatt 925
Db 1036 CACAGGACTTGGCATTCTTTCAGTTGCAAGGTTTATTCAAAATGGTGGTCTTCTTGGCACT 1095
QY 926 actacatgatcatgagagaacaattttggccgacatcatcagggtcttttcattgaa 985
Db 1096 ACTACATGATATCATGAGGAACAATTTTGGCCGAACCTGCTGGCGTCCCTTCATTGCA 1155
QY 986 cttagtcagattatgatgctctctcgtatgaatatgggttgggtcgaatgaacccaagtatg 1045
Db 1156 CTAGCTATGACTATGATGACCTCTAGATGAATATGGACTACCAAGGAACCAAAATATA 1215
QY 1046 ggcacttgagagactataaagctataaagctatcgaacccggtcttagtttccatcat 1105
Db 1216 GCCATTGGAATATATGCAAAAGCCATCAAGATGGCTGAGCCAGCTTTATTTGGCCACTG 1275
QY 1106 atgctgcggtgactagcttggaaatcaatcaagagctcatgtttatagatacaaaatctg 1165
Db 1276 ATGCTGCAGTGTCAAACTTGGAAATATATCAAGAGCTCATGTATACAGTCAAGATCTG 1335
QY 1166 gagcttgctgctcttttttccaaactatgactctagatatattcagtaaaagtcaactttc 1225
Db 1336 G---TTGTGCTGCATCTCTTGGCAATTTATGACACAAAATACCCCGTGAGAGTGACATTT 1392
QY 1226 agaatagggcactataatctgctcctcataggtccatcagcattcttcccgactgcaaaactg 1285
Db 1393 GGAATGAAGCAATATAACTTGCACCTTGGTCCATGAAGCATTTCTCTGATTGCAAAACTG 1452
QY 1286 ccgtttcaaacactcacaggttcaactctcaagctccagcataaagatgacgcctcag 1345
Db 1453 AAGTATTCAACTCTGAGGGTTGGC---CAGAGCCGCCCAACGAGAGATGACAC---CAG 1506
QY 1346 gtgttggtattctcttggcagtcatacaatgaagaacccctactctgctgatcagacgata 1405
Db 1507 TAGCTCACCTTTCTTGGCAAGCTTACATTGAAGATGTTGCCACATCTGCTGATGACAAATG 1566
QY 1406 cacttacagtaacgactatgggaacagaaaaacgcgcacagaagattcactcagactatc 1465
Db 1567 CATTTACATCAGTTGGACTTGTGTGAGCAAAATGAAGTCTCAGTTGGGATACACAGACTATT 1626
QY 1466 ttgtgtacatgacaaagttaaatatagcatctaatgaaggatttctaaagaacggaag 1525
Db 1627 TGTGTACATGACATATACAAATAGGTCTTAATGACAAATTTTAAAGACTGGAAGT 1686
QY 1526 atccttatctcactgttatgtccgctggtcattgtcttgcattgtttctcattggaacac 1585
Db 1687 ACCCCACTCTCAAAGTCGATTGAGTGGGCATGCCCTTGCATGTTTTCATCAATGGTCAAC 1746
QY 1586 taccgaagactgttatggtacattggtataatccaaaacttacatcacagtggcaacgtga 1645
Db 1747 TATCAGGATCTGCTTATGGAATCTTGGCGTTCGCCAAATTAAGATTTAATCAGGGTGTGA 1806
QY 1646 agttaagagctggtattcaagaatttctctgctcagttctcgttccgttcccggaacg 1705
Db 1807 AGCTGAGAGCTGGCAATAACAAGCTTGCTTTACTTTAGTTGTTTCAGTTGGTCTGGCGCAATG 1866
QY 1706 ttggcgtgcattatgatacgaatggaatgcaggagttcttaggtccagtcagcttgagcgtc 1765
Db 1867 TTGCTCTACACTTTGAGACATGGAAATAGTGGAGTTCTTGGGCGGCTCAGCTTGGCTGGTG 1926
QY 1766 tcaatgaagggtcaagaactggcgaacagaagaatggtcttacaaggttggtctgaaag 1825
Db 1927 TGAATTCGGGAACATGGGATATGACACCATGGCAATGGCACTTACCAAGATTTGGCATGAG 1986
QY 1826 gcgaactcgttaagcttccactccttaagtggaggttctctgttgtaatgggttcgagtt 1885
Db 1987 GTGAGATATGAGCTCCATACCGTCACTGGAAGTCTTCTTGTTGAATGGGTACAGGAT 2046

QY 1886 cactaatggtcctaaagcagccctgacttggtatacaaggtacatttaacgcgcctggag 1945
Db 2047 CACTATTGGCTCAATATCGACCCCTTACATGCTATAGAAGCTATTTAAATGCACCACAG 2106
QY 1946 gaaatgatccactagctatttagacatggcaagtattgggaaaaaggtcagatatgataaatg 2005
Db 2107 GAAATGCACCGTTCAGTCTTAGATATGGGTAGCATGGGAAGGTCAAAATGTGGATTAATG 2166
QY 2006 gtgaaggcgtaggtgcctcattggctgatacatgacagggcagcagcagcaaatgca 2065
Db 2167 GTACAGCATGCGACGCCCATTTGGCCGGCATATAAAGACACACGGGAGTTGTGGTGTGTT 2226
QY 2066 gttatgctggaaacttcaacgagaagaagtcggaactaaactgcggaacacctcttcaga 2125
Db 2227 ACTATGCTGGTACTTATACTGAAACAAATGTCGAACAAATTTGTGACAGCCCTCTCAGA 2286
QY 2126 gatggtacatgcttccacgactggtgctgaaacaaagtggaaacttggtagtagtattcg 2185
Db 2287 GATGTCACATGTTTCCCGCTCGTGGTTGAAGTCAAGTGGGAATTTATTTGGTGTGTTG 2346
QY 2186 aagaatgggaggttaattccaacaggaatttctctctagtcaggaga 2229
Db 2347 AAGAATGGGGTGGTATCCGACTAAGATTTCTTGGTAGCAAGA 2390
RESULT 14
AF004812
LOCUS AF004812 2202 bp mRNA linear PLN 22-JUN-1997
DEFINITION Mangifera indica beta-D-galactosidase (SP26) mRNA, partial cds.
ACCESSION AF004812
VERSION AF004812.1 GI:2209357
KEYWORDS
SOURCE mango.
ORGANISM Mangifera indica
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Rosidae; eusoids II; Sapindales; Anacardiaceae; Mangifera;
Paria-Arenas,S., Cruz-Hernandez,A., Gutierrez-Martinez,P. and
Gomez-Lim,M.A.
Cell wall metabolism in mango fruit: isolation of a cDNA clone to
beta-D-galactosidase and expression studies
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 2202)
AUTHORS Paria-Arenas,S., Cruz-Hernandez,A., Gutierrez-Martinez,P. and
Gomez-Lim,M.A.
Direct Submission
JOURNAL Submitted (19-MAY-1997) Genetic Engineering, CINVESTAV-Irapuato,
Apartado Postal 629, Irapuato, GTO 36500, Mexico
FEATURES
source
1..2202
/organism="Mangifera indica"
/strain="Manila"
/db_xref="taxon:29780"
/tissue_type="fruit"
<1..>2202
/gene="SP26"
214..>2202
/gene="SP26"
/EC_number="3.2.1.23"
/function="cell wall degradation; hydrolysis of
(1-4)-beta-D-galactans"
/codon_start=1
/product="beta-D-galactosidase"
/protein_id="AAB61470.1"
/db_xref="GI:2209358"
/translation="MGFEGSMTMLINMFMLLMLFSSWCFVEATVSDHKAILID
GQRILISGSIHYPRSTPQMKDVLQAKDGVDTQTVFVNGHPSPGPGYFEDRDY
LVRFILVQOAGLYVHLRIIGYVCAENFGGFPVWLKYVPGIERPNEPFFKAAMOKF
TEKIVSNMKAELFLFQGGPTILSOIENEFPGVWEIGAPCKAYTKWAQMAVGLDTG
VPVWCKQDDADPDVPVINTCNFGYCFENFVNQKNPKMWTENWGNWTFAGGPTQRP
EDVAFSVARFIQNGGSFVNVTMTHGTFNRTAGGTFIATSYDIDAPLEGLLEPRK
WGLRLDHLKALCESALVSTDPVTVTSLGNNQNEVHFNPKNKSGSCAAFLANYDTTSSAK

RESULT 15

LES012796 3010 bp mRNA linear PLN 07-JAN-1999
LOCUS Lycopersicon esculentum mRNA for ss-galactosidase, clone tegIA.
DEFINITION
ACCESSION AJ012796
VERSION
KEYWORDS beta-galactosidase; ss-1,4-exogalactanase; ss-galactosidase; tegIA gene.

SOURCE

ORGANISM Lycopersicon esculentum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; eusterids I; Solanales; Solanaceae; Solanum;
Lycopersicon.
1 (bases 1 to 3010)
De Silva, J.

REFERENCE

AUTHORS Direct Submission
JOURNAL Submitted (11-NOV-1998) de Silva J., Plant Sciences, Unilever
Research, Unilever Research, Colworth Laboratory, Sharnbrook,
Bedford, MK44 1LQ, UK
2 (bases 1 to 3010)
De Silva, J., Jarman, C., Strongitharm, B. and Gidley, M.
TITLE In vivo manipulation of plant cell walls
JOURNAL Unpublished

FEATURES

Location/Qualifiers
1..3010
/organism="Lycopersicon esculentum"
/cultivar="Money maker"
/db_xref="taxon:4081"
/clone="tegia"
/dev_stage="Fruit"
95..169
/gene="tegia"
95..2611
/gene="tegia"
/standard_name="beta-galactosidase"
/EC_number="3.2.1.23"
/function="ss-1,4-exogalactanase"
/codon_start=1
/evidence=experimental
/product="ss-galactosidase"
/protein_id="CAA10173.1"
/db_xref="GI:4138137"
/translation="MGCTLIILMLNVLVLLGSWFSFGTSVSYDHRALIVNGORRILI
SGSVHPRSTPEWMPGIIQKAKGGVDVIQTVFWNGHEPQQGKYIFEGRDYDLVKFIK
LVHAGLVHVRGPYACAEWNEFGFPVWLKVPVCGISPTDNGPEKAMOKFTAKIVN
MKAPRYETGGPIILSOIENYGPMEWELGAPCKSYAOWNAKMAVGLDTGVPWVC
KQDDAPDRIINACNGFYCDYSPNKAYKPKIWEANTAWFTGFGNPVYRPAEDLAFS
VAKFIQKGGSFINYMYHGGTNFGTFATSYDYDAPLDEYLLQRKPWGHLD
LHRAIKCEPALVSDPAVTALGHOEAHVFRSKAGSCAAF LANYDQHSFATVFNAR
HYNLPWISILPDCNVTNFTARIGASQAKMKTTPVSRGLPWQSFNEETSIEDSSF
TVVGLLEQINTROVSDYLVYSTDVKIDSRBKFLRGKWPWLTIMSAGHALHVFVNGQ
LAGTAYGSLKPKLTFKAVNLKAGVKNKISLSTAVGLPNIGPHFETWAGVLGPVSL
TGLDGRKDLTWKWSYKVLKGEALSLSGSSVSEWVEGSLVAQRP LWTYKSTF
NAPAGNDPLALDNTWKGQWVINGQSLGRYPWPGYKASNGACNAGWNEKKCLSN
CCEASQRYHVPVRLWYPTGLVLFEEWGGPHGISLVKREVASVCADINWOPQLV
NQMQASGVKDKPRLKHAHLSGASQKITSIKFASFGTPQGVGSGFREGSCHAFHSYD
AFERYCIGONSCSVPTPIEFGDDPCPHVMKKLSVEVICS"
95..2611
/gene="tegia"
170..2608
/gene="tegia"
/EC_number="3.2.1.23"
/function="ss-1,4-exogalactanase"
/product="ss-galactosidase"
883 a 529 c 732 g 866 t

gene

mat_peptide

BASE COUNT

ORIGIN
Query Match 45.9%; Score 1172.8; DB 8; Length 3010;
Best Local Similarity 71.0%; Pred No. 1.3e-216;
Matches 1568; Conservative 0; Mismatches 637; Indels 3; Gaps 1;

QY 80 tgttgtgtattagtattgtttattgttgatttttttttcttcttgtaagagctagtgttt 139
DB 117 TGTGTAATGTGTGTGGTGTGTGTTCATGGGTTTCTTCTGGAACAGCTTCGTCTT 176
QY 140 cttatgatcacagagctataatcataaataatgggaaagaaataattcttattctgggtcaa 199
DB 177 CATATCACCATAGGGCTATTATTGTAATGGACAAAGAAATACTATTCTTCGTCTCG 236
QY 200 ttcatattcaagaagacactccacagatgtggcctgatcttatcaaaaaggtcaaaagatg 259
DB 237 TTCAATTATCAAGACACTCTCGAGATGTGCCAGGTATTATTAAAGCTGGTGCAACAGCAG 296
QY 260 gaggcttagatgtattgaaacttatgttttcttggaatggacatgagccttctccctggaa 319
DB 297 GAGGTGTGATGTGATTCAGACTTATGTTTCTGGAATGGACATGAGCTCAACAAGGGA 356
QY 320 aataataattttgaagaagatgatgtctgttagattcattcaaaatggtacaaaagagcag 379
DB 357 AATATTATTATTGAAGGGAGATATGATTAGTGAAGTTTATTAAAGCTGGTGCAACAGCAG 416
QY 380 gactttatgcaattacgtattggccttagctgtgctgaatggaaactttgggggat 439
DB 417 GACTTTATGTCATCTTAGAGTTGGACCTTATGCTGTGCTGCTGAATGGAATTTTGGGGCT 476
QY 440 tccctgtttggctaaatatgtgctgtggtatggaatttagaacaacaatcgaccttta 499
DB 477 TTCTCTGTTTGGCTGAAATATGTTCCAGGTATCAGTTTCAGAACAGATTAATGGACCTTCA 536
QY 500 aggtgagctgaaggattgttcagaaaaatagtaacatgatgatgaagtcagaaaaattgt 559
DB 537 AGGCTGCAATGCAAAAATTTACTGCCAAGATTGTCAATATGATGAAGCGGAACGTTGT 596
QY 560 ttgaatctcaaggaggaccaataattatggtgccagatacaaaatgagtatggaccagttag 619
DB 597 ATGAACACTCAAGGGGGCCAAATAATTTTATCTCAGATTGAGAAATGAATGAGCCATCG 656
QY 620 aatgggaaattggtctcctgtaagcttatacaaaaatgggcagctcaaatggctgtag 679
DB 657 AATGGAACTGGGACACAGCAATAAATCTTACGCACAGTGGCGCCGCAAAATGGCTGTGG 716
QY 680 gtttgaactggtgtcccatgatcatgtgaagcaagagatgctctgatcctgtga 739
DB 717 GTCTTGACACTGGTGCCCATGGGTATGTGCAAGCAAGACATGCCCTGTATCTTATTA 776
QY 740 ttgatacttgtaaggcttactctcgaaaggttctcgtctcctaataagccttcaaaacct 799
DB 777 TAAATGCTTGCAATGGCTTCTACTGTGACTACTTTTCTCCAAACAAGGCTTATAACCAA 836
QY 800 aatgtggacagaagtatggactggtggtatcacgaaatcgggtggtccaaatctctcaa 859
DB 837 AGATATGGACTGAAGCCTGGACTGTCATGGTTTACTGGTTTGGAAATCCAGTTCCTTACC 896
QY 860 gaccagcgaagaacattgcaatttccagttgocaggtttgttcagaacaatggttcaatt 919
DB 897 GTCCCTGCTGAGACATGGGCATTTTCTGTGCAAAATTTATACAGAAGGAGGTTCTCTCA 956
QY 920 tcaattactacatgtatcatgatgaggaacaaatattggccgacatcatcagggcttttca 979
DB 957 TCAATTATTACATGTATCATGAGGAGNACAAACTTTGGACGGACTGCTGGTGGTCCATTTA 1016
QY 980 ttgcaactagctacgattatgatgtcctcctcgatgaatattgggtgtgtaagaaccaa 1039
DB 1017 TTGCTACTAGTTATGACTATGATGACACCTTGTGATGAATATGATGATTTTTCGACACCAA 1076
QY 1040 agtatggcacttgagagacttacataaaagctatcaagctatctgaacggcttagttt 1099
DB 1077 AATGGGTCACCTGAAAGATCTGCATAGAGCAATAAAGCTTTGTGAACACGAGTTAGTCT 1136
QY 1100 catcatatgctcggtgactagctcttggaagtaatacaagagctcatgtttatagatcaa 1159
DB 1137 CTGGAGATCCACTGTGACAGCAGCTTGGACACAGAGAGGCCCATGTGTTTAGGTGCA 1196
QY 1160 aatctggagcttgctgctgttttttatccaactatgactctagatatttcagtaaaagtc 1219

Db 2274 GTGTTTGTGCAGATATAAACGAATGGCAACACACAGTTGGTGAATTGGC 2321

Search completed: June 25, 2002, 06:21:18
Job time: 13608 sec

Db 1197 AGCTGGCTCTTGTGTCGCACTCTTGTAACTACGACCAACACACTCTTTTGTACTGTGT 1256
QY 1220 cctttcagaataggccatacaaatcctcctccatggtccatcagcattcttcccagactgca 1279
Db 1257 CATTTGCAACAGCGCATTAACAATGCGCCACCATGGTCAATCAGCANITCTTCCCGACTGCA 1316
QY 1280 aaactgcggtttacaacactgcacaggttaactcctcaagctcgagcataaagatgacgc 1339
Db 1317 AGAACACTGTATTTAATACAGCAGCGATCGGTCTCAAAAGTGTCTAGATGAAGATGACTC 1376
QY 1340 ctgcaggtggtgattctcttggcagtcatacaataaagaacagcctactgctgatgaca 1399
Db 1377 CAGTCAGCAGAGGATTGCCCTGGCAGTCATTCATGAAGAGACATCATCTTATCAAGACA 1436
QY 1400 gcgatacacattacagcgtgagactatgggaacagaaaaacgtcacaagagattcatcag 1459
Db 1437 G---TAGTTTACAGTGTGTGGCTATTTGGACAGACATATAATACAAGAGACGTGTCTG 1493
QY 1460 actatctgtgtacatgacaaatgtaaatatagcatctaatgaaggatttctctaaagaacg 1519
Db 1494 ATTATTTGTGGTATTCAACAGATGTCAAGATTGATTCAGAGAAAAAGTTTTCAGAGGCG 1553
QY 1520 gaaagatccttatctcaactgttatccgctggtcatgtcttgcatgtttttogtcaatg 1579
Db 1554 GAAATGTGGCTTTGGCTTACGATCATGTCTAGCTGGGATGCAATTCATGTTTGTGAATG 1613
QY 1580 gaaactatcagaaactgtttatggtacattggtataatccaaacitacatacagtgga 1639
Db 1614 GTCAAATTAGCAGGAACGTGCATATGGAAGTTAGAAAAACCGAAACCTAACCTTCAGTAAAG 1673
QY 1640 acgtgaagttgaagctggtattacaagaatttctctgctcagtggttccgttgggtctcc 1699
Db 1674 CCGTAAATCTGACGAGGAGGTTTAAACAAGATTCTCTACTGAGCAUATGCTGTGGGCTTC 1733
QY 1700 cgaacgttggcggtgattatgatacatggaatgcagaggttcttagtccagtcacgttga 1759
Db 1734 CGAATATCGGCCACACATTTTGAGACATGGAATGCTGGTGTCTTGGGCCACTCTCACTAA 1793
QY 1760 gcggtctcaatgaagggtcaagaaacttggcgaacagaaatggtctctacaaggttggctc 1819
Db 1794 CTGGTCTTGACGAGGGGAAAGAGATTAAACATGGCAGAAATGGTCTTACAAGTTGGTC 1853
QY 1820 tgaagcgaaactggttaagtcttcaactccttaagtgggagttcttctgttgaatgggttc 1879
Db 1854 TAAAAGGAGAAGCCTTGAGCCTCCATTCACTCAGTGTGTAGCTCGTCAGTTGAGTGGGTG 1913
QY 1880 gaggttcaataatgggtcaaaagcagccctgacttgggtacaagggtacatttaacgcgc 1939
Db 1914 AGGGTTCTTTAGTGGGTCAGAGACGCCACTACATGGTACAAGAGCACTTTTAAATGCTC 1973
QY 1940 ctggaggaaatgatccactagctcttagacatggaagtagtggaagaggtcagatatgga 1999
Db 1974 CAGCTGGAATATGCTTTGGCTTTAGACTTGAATACCATGGGCAAGGACAAAGTGTGGA 2033
QY 2000 taaatggtgaagcgtaggtgcgcatctggcctgtgatacatagcacaaggcgactgcagca 2059
Db 2034 TAAATGGTCAAAAGCCTCGGAGCGTATTGGCCTGGATATAAAGCATCTGTTAACTGCGGTG 2093
QY 2060 atgcggttatgctgaaagttcaacgagaagtcgcggaactaaactgcggacaacctt 2119
Db 2094 CCTGTAACTATGAGGCTGGTGTAAATGAGAAAAAATGCCTAAGTAACCTGTGGAGAGGCTT 2153
QY 2120 ctccagatggttaccatgctccacgtcgtgctgaaacacagtggaactgtttagtag 2179
Db 2154 CACAACGATGGTATCATGTTCCCGTCTTGGCTGTATCTCTACTGGAAATTTGTTAGTTC 2213
QY 2180 tattcgaagaatgggaggttaatcccaacaggaatttctctagtcaggagatcaagataaa 2239
Db 2214 TATTTGAGGAATGGGAGGAGAGCCCTCATGGAATCTCTTTGGTAAAAAGAGAGATTGCAA 2273
QY 2240 gaactcgaaagtaaaactgttgcagtaactatggtgcttgaattcgc 2287

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2002, 20:37:40 ; Search time 34.31 Seconds
(without alignments)
3650.490 Million cell updates/sec

Title: US-09-701-868-11
Perfect score: 3924
Sequence: 1 MERTNVLLLVICLLDFSS.....VVFEWGGNPTGISLVRRSR 724

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues
Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPATREMBL09

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phage.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_rviris.*
- 16: sp_bacteriap.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	3924	100.0	724	10	O81100 lycopersico
2	3909	99.6	724	10	Q9T0P6 lycopersico
3	3632	92.6	724	10	Q93XM4 capsicum an
4	3063.5	78.1	721	10	Q9ZP30 carica papa
5	3055	77.9	731	10	Q9AYS1 pyrus pyrif
6	3028.5	77.2	721	10	Q9M5J4 phaseolus a
7	2992.5	76.3	730	10	O65736 cicer ariet
8	2948.5	75.1	723	10	O82670 cicer ariet
9	2941	74.9	838	10	Q9ZP11 lycopersico
10	2928.5	74.6	847	10	Q9SCW1 arabidopsis
11	2915	74.3	843	10	Q93X58 fragaria an
12	2906	74.1	730	10	Q9ZP17 lupinus ang
13	2901.5	73.9	722	10	Q93X56 fragaria an
14	2848.5	72.6	728	10	Q9SCV0 arabidopsis
15	2838	72.3	729	10	Q9SZI5 arabidopsis
16	2809.5	71.6	724	10	Q93Y27 arabidopsis

17	2808.5	71.6	724	10	Q9SCV8	Q9scv8 arabidopsis
18	2807	71.5	727	10	Q9SCW0	Q9scw0 arabidopsis
19	2797	71.3	727	10	Q9LPA6	Q9lfa6 arabidopsis
20	2736	69.7	663	10	O04976	O04976 mangifera l
21	2687.5	68.5	854	10	Q94B17	Q94b17 vitis vinif
22	2653.5	67.6	745	10	Q92RV9	Q92rv9 cicer ariet
23	2636.5	67.2	845	10	Q9LIS9	Q9lls9 lycopersico
24	2617.5	66.7	739	10	Q9M5J3	Q9m5j3 phaseolus a
25	2608	66.5	853	10	O23243	O23243 arabidopsis
26	2608	66.5	856	10	Q9SCV9	Q9scv9 arabidopsis
27	2572	65.5	853	10	Q42150	Q42150 arabidopsis
28	2560	65.2	732	10	Q9MAJ7	Q9maj7 arabidopsis
29	2560	65.2	732	10	Q9SCV7	Q9scv7 arabidopsis
30	2556	65.1	732	10	Q93263	Q93263 arabidopsis
31	2369.5	60.4	707	10	O65761	O65761 cicer ariet
32	2261	57.6	852	10	Q9SCV4	Q9scv4 arabidopsis
33	2259.5	57.6	839	10	Q9SK11	Q9sk11 arabidopsis
34	2219	56.5	852	10	Q9LLT0	Q9llt0 lycopersico
35	2177	55.5	840	10	Q93X57	Q93x57 fragaria an
36	2103.5	53.6	741	10	Q9FN08	Q9fn08 arabidopsis
37	2097.5	53.5	741	10	Q9SCV2	Q9scv2 arabidopsis
38	2049.5	52.2	887	10	Q9SCV3	Q9scv3 arabidopsis
39	2035.5	51.9	895	10	O48836	O48836 arabidopsis
40	1843	47.0	788	10	Q9SCV5	Q9scv5 arabidopsis
41	1811	46.2	718	10	Q9SCV6	Q9scv6 arabidopsis
42	1808	46.1	718	10	Q9FFN4	Q9ffn4 arabidopsis
43	1786	45.5	779	10	Q9C6W4	Q9c6w4 arabidopsis
44	1682	42.9	715	10	Q9FSF9	Q9fsf9 nicotiana t
45	1649	42.0	808	10	Q94728	Q94728 oryza sativ

ALIGNMENTS

RESULT 1
O81100 PRELIMINARY; PRT; 724 AA.
ID O81100;
AC /01-NOV-1998/ (Tremblrel. 08, Created)
DT /01-NOV-1998/ (Tremblrel. 08, Last sequence update)
DT /01-DEC-2001/ (Tremblrel. 19, Last annotation update)
DE BETA-GALACTOSIDASE PRECURSOR (EC 3.2.1.23) (LACTASE).
GN BGAL4.
OS Lycopersicon esculentum (Tomato).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
OX NCBI_TaxID=4081;
RN [1]
RC SEQUENCE FROM N.A.
RP STRAIN=RUTGERS; TISSUE=Tomato FRUIT;
RX MEDLINE=98289087; PubMed=9625694;
RA Smith D.L., Starrett D.A., Gross K.C.;
RT "A gene coding for tomato fruit beta-galactosidase II is expressed during fruit ripening. Cloning, characterization, and expression pattern."
RT Plant Physiol. 117:417-423(1998).
RL -1- CATALYTIC ACTIVITY: HYDROLYSIS OF TERMINAL, NON-REDUCING BETA-D-
CC -1- GALACTOSE RESIDUES IN BETA-D-GALACTOSIDES.
CC -1- SIMILARITY: BELONGS TO FAMILY 35 OF GLYCOSYL HYDROLASES.
DR EMBL; AF020330; AAC25984.1; -
DR InterPro; IPR001944; Glyco_hydro_35.
DR Pfam; PF013101; Glyco_hydro_35; 1.
DR PRINTS; PR00742; GLYHDRLASE35.
DR PROSITE; PS01182; GLYCOSYL_HYDROL_F35; 1.
KW Glycosidase; Hydrolase; Trans it peptide.
FT TRANSIT 1 23 POTENTIAL.
FT ACT_SITE 181 181 PROTON DONOR (BY SIMILARITY).
SQ SEQUENCE 724 AA; 80513 MW; 6F369499E25BDB2C CRC64;

Query Match 100.0%; Score 3924; DB 10; Length 724;
Best Local Similarity 100.0%; Pred. No. 1.1e-256;

[illegible]

3

```
Q93XM4
ID Q93XM4 PRELIMINARY; PRT; 724 AA.
AC Q93XM4;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE BETA-GALACTOSIDASE.
GN BGI.
OS Capsicum annuum (Bell pepper).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; euasterids I; Solanales; Solanaceae; Capsicum.
OX NCBI_TaxID=4072;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. MC11; TISSUE=RIPENING FRUIT PERICARP;
RA Zainal-Abidin W., Zainal Z., Ali Z.M., Lazan H.;
RT "A full length cDNA sequence of Capsicum annuum L. beta
galactosidase.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY029226; AAK40304.1; -.
SQ SEQUENCE 724 AA; 80176 MW; DE9BD5B6E09D4DC9 CRC64;

Query Match 92.6%; Score 3632; DB 10; Length 724;
Best Local Similarity 91.2%; Pred. No. 5.7e-237;
Matches 662; Conservative 38; Mismatches 22; Indels 4; Gaps 2;

Qy 1 MLRTN--VLLLVICLLDFSSVKASVSYDDRAIIINGKRKILISGSIHYPRSTPQWMPD 58
Db 1 MKMSNNVLLVVLVTCSDLL--VKANSYDDRAIVINGKRKILISGSIHYPRSTPQWMPD 58
Qy 59 LIQAKDGGDLVITYYFVWNGHEPSPKYNFEGRYDLVRFIKMVORAGLYVNLRIQPVYC 118
Db 59 LIEAKDGGDLVITYYFVWNGHEPSPKYNFEGRYDLVRFIKLVQAGLYVNLRIQPVYC 118
Qy 119 AENFNGFPVWLKYVPGMEFTNNQPFKAMQGVQKIVNMKSENLFESQGGPIIAQI 178
Db 119 AENFNGFLPVWLKYVSGMEFTNDQPFKAMQGVQKIVSMKSEKLFEPQGGPIIAQI 178
Qy 179 ENYGVPEWELGAPKAYTKWAAQMAVGLKTGVPWIMCKQEDADPDVIDTCNGFYCEGR 238
Db 179 ENYGVPEWELGAPKAYTKWAAQMAVGLKTGVPWIMCKQEDADPDVIDTCNGFYCEGR 238
Qy 239 PNKPKPKMTEVMTGWYTKFGGPIQORPAEDIAFSVAREFVQNGSFNFYMYHGGTFNG 298
Db 239 PNKPKPKMTEVMTGWYTKFGGPIQORPAEDIAFSVAREFVQNGSFNFYMYHGGTFNG 298
Qy 299 RTSSGLFIATSYDYDAPLDEYGLLNEPKYGHRLDLHKAIKLSEPALVSSYAATVSLGSNQ 358
Db 299 RTSSGLFIATSYDYDAPLDEYGLLNEPKYGHRLDLHKAIKOCEPALVSSYPTVTSLSNQ 358
Qy 359 EAHVYRSKSGACAAFLSNYDSRISVKYTFQNPYNLPWISILPDCKTAVYNTAQVNSQ 418
Db 359 EAHVYRSKSGACAAFLSNYDAKYSVRVSFQNLPYDLTPWISILPDCKTAVYNTAKVSSQ 418
Qy 419 SSSIKMTFPAAGGLSWQSYNETPTADSDTLTANGLMEQKNVTRDSDYLWYMTNVTAS 478
Db 419 GSSIKMTFPAAGGLSWQSYNETPTADSDTLRANGLMEQKNVTRDSDYLWYMTDNVTAS 478
Qy 479 NEGLKNGKDPYLVMSAGHVLHVFNKGLSGTYGTLNDPKLTYSGNVKLKLAGINKISL 538
Db 479 NEGLKNGKDPYLVMSAGHVLHVFNKGLSGTYGTLNDPKLTYSGNVKLKLAGINKISL 538
Qy 539 LSVSVGLPNVGVHVDTNAGVLGVPVTLISGLNEGRNLAKQKWSYKVGKLSLSLSLSG 598
Db 539 LSVSVGLPNVGVHVDTNAGVLGVPVTLISGLNEGRNLAKQKWSYKVGKLSLSLSLSG 598
Qy 599 SSSVWVVRGSLMAOKPLTWYKATFNAPGGNDPLALDMSMGKGOIWTNGSVGRHWPY 658
Db 599 SSSVWVVRGSLMAOKPLTWYKATFNAPGGNDPLALDMSMGKGOIWTNGSVGRHWPY 658
Qy 659 IAQGDCSKCSYAGTFNEKKCTQNCQPSQRWYHVPWSLKPDSGNLLVVFEEWGGNPTGIS 718
Db 659 IAQGDCSKCSYAGTFNEKKCTQNCQPSQRWYHVPWSLKPDSGNLLVVFEEWGGNPTGIS 718

Db 659 AAQGDCSKCSYAGTFNEKKCTQNCQPSQRWYHVPWSLKPDSGNLLVVFEEWGGNPTGIS 718
Qy 719 LVRRSR 724
Db 719 LVRRSR 724

RESULT 4
Q92P30 PRELIMINARY; PRT; 721 AA.
AC Q92P30;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE BETA-GALACTOSIDASE PRECURSOR (EC 3.2.1.23) (LACTASE).
OS Carica papaya (Papaya).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Caricaceae; Carica.
OX NCBI_TaxID=3649;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. EKSOTIKA; TISSUE=MESOCARP;
RA Othman R., Choo T.S., Ali Z.M., Zainal Z., Lazan H.;
RT "A full-length beta-galactosidase cDNA sequence from ripening
papaya.";
RL Plant Physiol. 118:1102-1102(1998).
CC -1- CATALYTIC ACTIVITY: HYDROLYSIS OF TERMINAL, NON-REDUCING BETA-D-
CC -1- GALACTOSE RESIDUES IN BETA-D-GALACTOSIDES.
CC -1- SIMILARITY: BELONGS TO FAMILY 35 OF GLYCOSYL HYDROLASES.
DR EMBL; AF064786; AAC77377.1; -.
DR InterPro; IPR001944; Glyco_hydro_35.
DR Pfam; PF01301; Glyco_hydro_35; 1.
DR PRINTS; PR00742; GLYHDLASE35.
DR PROSITE; PS01182; GLYCOSYL-HYDROL_F35; 1.
KW Glycosidase; Hydrolase; Signal.
FT SIGNAL 1 18 POTENTIAL.
FT CHAIN 19 721 BETA GALACTOSIDASE.
SQ SEQUENCE 721 AA; 80985 MW; B41BF5E16C299617 CRC64;

Query Match 78.1%; Score 3063.5; DB 10; Length 721;
Best Local Similarity 75.2%; Pred. No. 1.4e-198;
Matches 543; Conservative 86; Mismatches 90; Indels 3; Gaps 2;

Qy 1 MLRTNVLVLLVLCILDFSSVKASVSYDDRAIIINGKRKILISGSIHYPRSTPQWMPDLI 60
Db 1 MLRTNVLVLLVLCILDFSSVKASVSYDDRAIIINGKRKILISGSIHYPRSTPQWMPDLI 60
Qy 61 QKAKDGGDLVITYYFVWNGHEPSPKYNFEGRYDLVRFIKMVORAGLYVNLRIQPVYCAE 120
Db 61 QKAKDGGDLVITYYFVWNGHEPSPKYNFEGRYDLVRFIKLVQAGLYVNLRIQPVYCAE 120
Qy 59 QNAKEGGLDVITVYFVWNGHEPSPKYNFEGRYDLVRFIKLVQAGLYVNLRIQPVYCAE 118
Db 59 QNAKEGGLDVITVYFVWNGHEPSPKYNFEGRYDLVRFIKLVQAGLYVNLRIQPVYCAE 118
Qy 121 WNFEGFPVWLKYVPGMEFTNNQPFKAMQGVQKIVNMKSENLFESQGGPIIAQIN 180
Db 121 WNFEGFPVWLKYVPGMEFTNNQPFKAMQGVQKIVNMKSENLFESQGGPIIAQIN 180
Qy 119 WNFEGFPVWLKYVPGIOPRTDNGPFAQOMQKFTKIVNMKAEKLFEPQGGPII 178
Db 119 WNFEGFPVWLKYVPGIOPRTDNGPFAQOMQKFTKIVNMKAEKLFEPQGGPII 178
Qy 181 EYGPVWEIGAPKAYTKWAAQMAVGLKTGVPWIMCKQEDADPDVIDTCNGFYCEGRPN 240
Db 181 EYGPVWEIGAPKAYTKWAAQMAVGLKTGVPWIMCKQEDADPDVIDTCNGFYCEGRPN 240
Qy 179 EYGPVWEIGAPKAYTKWAAQMAVGLKTGVPWIMCKQEDADPDVIDTCNGFYCEGRPN 238
Db 179 EYGPVWEIGAPKAYTKWAAQMAVGLKTGVPWIMCKQEDADPDVIDTCNGFYCEGRPN 238
Qy 241 KPYKPKMTEVMTGWYTKFGGPIQORPAEDIAFSVAREFVQNGSFNFYMYHGGTFNGRT 300
Db 241 KPYKPKMTEVMTGWYTKFGGPIQORPAEDIAFSVAREFVQNGSFNFYMYHGGTFNGRT 300
Qy 239 ANYKPKMTEVMTGWYTKFGGPIQORPAEDIAFSVAREFVQNGSFNFYMYHGGTFNGRT 298
Db 239 ANYKPKMTEVMTGWYTKFGGPIQORPAEDIAFSVAREFVQNGSFNFYMYHGGTFNGRT 298
Qy 301 SSGFLFIATSYDYDAPLDEYGLLNEPKYGHRLDLHKAIKLSEPALVSSYAATVSLGSNOEA 360
Db 301 SSGFLFIATSYDYDAPLDEYGLLNEPKYGHRLDLHKAIKLSEPALVSSYAATVSLGSNOEA 360
Qy 299 AGGFIATSYDYDAPLDEYGLLNEPKYGHRLDLHKAIKLSEPALVSSYAATVSLGSNOEA 358
Db 299 AGGFIATSYDYDAPLDEYGLLNEPKYGHRLDLHKAIKLSEPALVSSYAATVSLGSNOEA 358
Qy 361 HVIYRSKSGACAAFLSNYDSRISVKYTFQNPYNLPWISILPDCKTAVYNTAQVNSOSS 420
Db 361 HVIYRSKSGACAAFLSNYDSRISVKYTFQNPYNLPWISILPDCKTAVYNTAQVNSOSS 420
Qy 359 HVFWTKT--SCAAFLANYDLKYSVRVTFQNLPPWSVILPDCKTAVYNTAQVNSOSS 417
Db 359 HVFWTKT--SCAAFLANYDLKYSVRVTFQNLPPWSVILPDCKTAVYNTAQVNSOSS 417
```



```

Db 10 VLMLFFWVCVTSVYTHDKAIVIDGKRILLISGSIHYPRSTPQMPPDLIOAKDGG 69
QY 70 VIETVFWNGHEPSPGKYNEGRDYLVRFTKMWQAGLYVNLRIYVCAEWNEFGFPVW 129
Db 70 VIETVFWNGHEPSPGKYNEGRDYLVRFTKMWQAGLYVNLRIYVCAEWNEFGFPVW 129
QY 130 LKYPVMEFRTNQPFKVMQGFVQKIVNMKSENLFESQGGPIIMAOIENEYGPVWEI 189
Db 130 LKYPVMEFRTNQPFKVMQGFVQKIVNMKSENLFESQGGPIIMAOIENEYGPVWEI 189
QY 190 GAFCKATYKAAQMAVGLKTVGPWIMCKQEDADPPVIDTCNGFCYCEFRPNKPKMWT 249
Db 190 GAFCKATYKAAQMAVGLKTVGPWIMCKQEDADPPVIDTCNGFCYCEFRPNKPKMWT 249
QY 250 EVMTGWTKFGGPIQORPAEDIAFSAVAREVQNGSFYNYMHGCTNFGRTSSGLFIATS 309
Db 250 EVMTGWTKFGGPIQORPAEDIAFSAVAREVQNGSFYNYMHGCTNFGRTSSGLFIATS 309
QY 310 YDYPADLDEYGLLNEPKYGHRLDLHKAIKLSEPALVSSAAVTSLSGNSQEAHVYRSKGA 369
Db 310 YDYPADLDEYGLLNEPKYGHRLDLHKAIKLSEPALVSSAAVTSLSGNSQEAHVYRSKGA 369
QY 370 CAAPLSNYSRYSVKVTFQRPNPLPWSISILPCKTAVNTAQNVSQSSIKMTFAGG 429
Db 370 CAAPLSNYSRYSVKVTFQRPNPLPWSISILPCKTAVNTAQNVSQSSIKMTFAGG 429
QY 430 GLWSQSYNEETPADDDSTLTANGLEQKNVTRDSSDYLWMTNVTASNEGFLKNCKDP 489
Db 430 GLWSQSYNEETPADDDSTLTANGLEQKNVTRDSSDYLWMTNVTASNEGFLKNCKDP 489
QY 490 YLVMSAGHVLHVFNGLSGVYGTLDNPKLTVSGNVKLAGINKISLSVSVGLPNVG 549
Db 490 YLVMSAGHVLHVFNGLSGVYGTLDNPKLTVSGNVKLAGINKISLSVSVGLPNVG 549
QY 550 VHDYTNAGVLGPTVTLGSLNEGRNLAKQKWSYKVLGKESLSLHSLSGSSVVEWVRS 609
Db 550 VHDYTNAGVLGPTVTLGSLNEGRNLAKQKWSYKVLGKESLSLHSLSGSSVVEWVRS 609
QY 610 MAOKQPLTWKATFNAPGGNDPLALDMSMGKGOIWTNGEGVGRHPGYIAQGDCSKCY 669
Db 610 MAOKQPLTWKATFNAPGGNDPLALDMSMGKGOIWTNGEGVGRHPGYIAQGDCSKCY 669
QY 670 AGTFNEKKQTCNQGPSQRYHVPRLKPSGNLLVVFEEWGGNPTGISLVRSS 723
Db 670 AGTFNEKKQTCNQGPSQRYHVPRLKPSGNLLVVFEEWGGNPTGISLVRSS 723
QY 671 AGTYDQKCRKTCNQGPSQRYHVPRLKPSGNLLVVFEEWGGNPTGISLVRSS 723
Db 671 AGTYDQKCRKTCNQGPSQRYHVPRLKPSGNLLVVFEEWGGNPTGISLVRSS 723
RESULT 7
ID O65736 PRELIMINARY; PRT; 730 AA.
AC O65736;
[1]
DT 01-NOV-1998 (Tremblrel. 07, Created)
DE 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DE BETA-GALACTOSIDASE (EC 3.2.1.23) (LACTASE).
OS Cicer arietinum (Chickpea) (Garbanzo).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids I; Fabales; Fabaceae; Papilionoideae; Cicer.
OX NCBI_TaxID=3827;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV, CASTELLANA; TISSUE=ETIOLATED EPICOTYLS;
RA Esteban R., Dopico B., Labrador E.;
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
CC -1- CATALYTIC ACTIVITY: HYDROLYSIS OF TERMINAL, NON-REDUCING BETA-D-
CC GALACTOSE RESIDUES IN BETA-D-GALACTOSIDES.
CC -1- SIMILARITY: BELONGS TO FAMILY 35 OF GLYCOSYL HYDROLASES.
DR EMBL: AJ005042; GenBank: AF06309.1;
DR InterPro: IPR001944; Glyco_hydro_35.
DR Pfam: PF01301; Glyco_hydro_35; 1.
DR PRINTS: PR00742; GLHYDRLASE35.

```

```

DR PROSITE; PS01182; GLYCOSYL_HYDROL_F35; 1.
KW Glycosidase; Hydrolase.
SQ SEQUENCE 730 AA; 81300 MW; F79252ABFBD1E6C3 CRC64;

Query Match 76.3%; Score 2992.5; DB 10; Length 730;
Best Local Similarity 73.9%; Pred. No. 8.8e-194;
Matches 530; Conservative 88; Mismatches 96; Indels 3; Gaps 2;

QY 8 LLLVICLLDFSSKASVSDDRAIIINGKRILLISGSIHYPRSTPQMPPDLIOAKDGG 67
Db 16 LVLELCL--FVSVTASVYDHLKAIINGQRILLISGSIHYPRSTPQMPPDLIOAKDGG 73
QY 68 LDVIETVFWNGHEPSPGKYNEGRDYLVRFTKMWQAGLYVNLRIYVCAEWNEFGFP 127
Db 74 VDIQTVFWNGHEPSPGKYNEGRDYLVRFTKMWQAGLYVNLRIYVCAEWNEFGFP 133
QY 128 VLVKYPVMEFRTNQPFKVMQGFVQKIVNMKSENLFESQGGPIIMAOIENEYGPVWE 187
Db 134 VLVKYPVMEFRTNQPFKVMQGFVQKIVNMKSENLFESQGGPIIMAOIENEYGPVWE 193
QY 188 EIGAPGRAYTKAAQMAVGLKTVGPWIMCKQEDADPPVIDTCNGFCYCEFRPNKPKM 247
Db 194 EIGAPGRAYTKAAQMAVGLKTVGPWIMCKQEDADPPVIDTCNGFCYCEFRPNKPKM 253
QY 248 WTEVMTGWTKFGGPIQORPAEDIAFSAVAREVQNGSFYNYMHGCTNFGRTSSGLFI 307
Db 254 WTEVMTGWTKFGGPIQORPAEDIAFSAVAREVQNGSFYNYMHGCTNFGRTSSGLFI 313
QY 308 TSQYDAPLDEYGLLNEPKYGHRLDLHKAIKLSEPALVSSAAVTSLSGNSQEAHVYRS 367
Db 314 TSQYDAPLDEYGLLNEPKYGHRLDLHKAIKLSEPALVSSAAVTSLSGNSQEAHVYRS 373
QY 368 GACAAFLSNYSRYSVKVTFQRPNPLPWSISILPCKTAVNTAQNVSQSSIKMTFAG 426
Db 374 GACAAFLSNYSRYSVKVTFQRPNPLPWSISILPCKTAVNTAQNVSQSSIKMTFAG 433
QY 427 AGGSLWSQSYNEETPADDDSTLTANGLEQKNVTRDSSDYLWMTNVTASNEGFLKN 486
Db 434 VSSAFDMSQSYNEETPADDDSTLTANGLEQKNVTRDSSDYLWMTNVTASNEGFLKN 493
QY 487 KDPVLTVMAGHVLHVFNGLSGVYGTLDNPKLTVSGNVKLAGINKISLSVSVGLP 546
Db 494 QYVLTVMAGHVLHVFNGLSGVYGTLDNPKLTVSGNVKLAGINKISLSVSVGLP 553
QY 547 NVGVHDTNAGVLGPTVTLGSLNEGRNLAKQKWSYKVLGKESLSLHSLSGSSVVEW 606
Db 554 NVGVHDTNAGVLGPTVTLGSLNEGRNLAKQKWSYKVLGKESLSLHSLSGSSVVEW 613
QY 607 GSLMAQKQPLTWKATFNAPGGNDPLALDMSMGKGOIWTNGEGVGRHPGYIAQGDCSK 666
Db 614 GSSLVKKQPLTWKATFNAPGGNDPLALDMSMGKGOIWTNGEGVGRHPGYIAQGDCSK 673
QY 667 CSYAGTFNEKKQTCNQGPSQRYHVPRLKPSGNLLVVFEEWGGNPTGISLVRSS 723
Db 674 CNYAGTFNEKKQTCNQGPSQRYHVPRLKPSGNLLVVFEEWGGNPTGISLVRSS 730

RESULT 8
ID O82670 PRELIMINARY; PRT; 723 AA.
AC O82670;
[1]
DT 01-NOV-1998 (Tremblrel. 08, Created)
DE 01-NOV-1998 (Tremblrel. 08, Last sequence update)
DE 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE BETA-GALACTOSIDASE (EC 3.2.1.23) (LACTASE).
OS Cicer arietinum (Chickpea) (Garbanzo).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids I; Fabales; Fabaceae; Papilionoideae; Cicer.
OX NCBI_TaxID=3827;
RN [1]
RP SEQUENCE FROM N.A.

```

```
RC STRAIN=CV. CASTELLANA; TISSUE=ETIOLATED EPICOTYLS;
RA Duplico B., Esteban R., Labrador E.;
RT "The fourth beta-galactosidase clone found in Cicer epicotyls.";
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases
CC -1- CATALYTIC ACTIVITY: HYDROLYSIS OF TERMINAL, NON-REDUCING BETA-D-
CC GALACTOSE RESIDUES IN BETA-D-GALACTOSIDES.
CC -1- SIMILARITY: BELONGS TO FAMILY 35 OF GLYCOSYL HYDROLASES.
DR EMBL; AJ011010; CA09457.1;
DR InterPro: IPR001944; Glyco_hydro_35.
DR Pfam: PF01301; Glyco_hydro_35; 1.
DR PRINTS: PR00742; GLHYDRLASE35.
DR PROSITE: PS01182; GLYCOSYL_HYDROL_F35; 1.
KW Glycosidase; Hydrolase.
SQ SEQUENCE 723 AA; 80655 MW; 0585760D631B293A CRC64;

Query Match 75.18; Score 2948.5; DB 10; Length 723;
Best Local Similarity 72.68; Pred. No. 8.1e-191;
Matches 522; Conservative 88; Mismatches 106; Indels 3; Gaps 2;

QY 5 NVL-LLLVICLLDFSSVKASVYDDRAIIINGKRKILISGSIHYPRSTPOMWPDLIQKA 63
DB 6 NLGLMLLLCF--WCAVTASVYDHTIVIDGQRRLISGSIHYPRSTPEWPFALFQKA 63
QY 64 KGGDLVETIYVFWNGHEPSPGKYNFEGRYDLVRFIKMVRAGLYVNLRIQGYVCAEWNF 123
DB 64 KEGGLDVIQTIVFWNGHEPSPGKYFEDRDLVRFIKLAQAGLYVHLRIQGYVCAEWNF 123
QY 124 GGFVPLKYYVPGMEFRINQPFKVMQGFVOKIYNNMKSENLFESQGGPIIQAQIENYG 183
DB 124 GGFVPLKYYVPGISFRDINPEFKAAMQKFTTKIYVSMKAENLFQNGGPIIQAQIENYG 183
QY 184 PVEVEIGAPGKAYTKAAQMAVGLTKTGVPTMCKOEADPDVIDTCNGFCYCEGFRPNKY 243
DB 184 PVEVNIGAPGKAYTNAAQMAVGLDTGVPTMCKOEADPDVIDTCNGYCENTFPKNY 243
QY 244 KPKMWTETVGTWYTKFGGPIPORPAEDIAFVARFVQNGSFFNYMYHGTNFGRTSSG 303
DB 244 KPKMWTENSGWYTFDGNIAICRPVEDLAYSVARFIQNRGSFVNYMYHGTNFGRTSSG 303
QY 304 LFIATSYDYDAPDEYGLLNEPKYCHLRLHKAIKLSEPALVSSYAAVTSLSGNSQEAHY 363
DB 304 LFIATSYDYDAPIDEYGLTNEPKWHLRLHKAIKQCEPALVSDVPTIITSLGNKLEAHY 363
QY 364 RSKSGACAALNSYDSRVSVKVTQNRPNLPWPSISILPDKCTAVNTAQNVSQSSSIK 423
DB 364 STGTSVCAAFLANDYTKSAATVTGNGKYDLPWNSVILPDKCTDVNTAKVGAQSSQKT 423
QY 424 MTPAGGGLSWQSYNEETPTADSDTLTANGLWEOKNVTROSSDYLVWNTVNTVNIASNEGFL 483
DB 424 MISTNSTFDWQSYTEEPAFSDESDSITAELWEQINVTROSSDYLVWNTVNTVNIASNEGFL 483
QY 484 KNGKDPYLTYNMAGHVLHVFNGLSGTVYGLTNDPKLTYSGNVKLRAGINKILLSVSV 543
DB 484 KNGQYPIILNMSAGHVLHVFNGLSGTVYGLVNDPKLTFNSNVNLTVGNKILLSVAV 543
QY 544 GLPNVGLHFEFWNVGVLGPTVTLKGLNEGRDLSWQKSYKVLKESLSLHTITGSSVD 603
DB 544 GLPNVGLHFEFWNVGVLGPTVTLKGLNEGRDLSWQKSYKVLKESLSLHTITGSSVD 603
QY 604 WVRGSLMAQKPLTWYKATFNAPGNPDALDMSMGKQIWIINGEGRHWPGYIAQGD 663
DB 604 WTQGSLLAKKPLTWYKATFNAPGNPDPLGDMSSMGKEIWNQDSTGRHWPGYIAHGS 663
QY 664 CSKCSYAGTFNEKCOINCGOPQSRWVHVPKSLKPSGNLLVVEEWGCPNTGISLYVR 722
DB 664 CGDCDYAGTFNTKRCNTCGNPTQWYHIPKSWLNPTGNVLVLEEWGDPGSGISLLKR 722
```

RESULT 9

Q9ZP11

• ID Q9ZP11 PRELIMINARY; PRT; 838 AA.

AC Q9ZP11;

```
DT 01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE BETA-GALACTOSIDASE PRECURSOR (EC 3.2.1.23) (LACTASE).
GN TEGIA.
OS Lycopersicon esculentum (Tomato).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
OX NCBI_TaxID=4081;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. MONEY MAKER;
RA De Silva J., Jarman C., Strongitharm B., Gidley M.;
RT "In vivo manipulation of plant cell walls.";
RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases
CC -1- CATALYTIC ACTIVITY: HYDROLYSIS OF TERMINAL, NON-REDUCING BETA-D-
CC GALACTOSE RESIDUES IN BETA-D-GALACTOSIDES.
CC -1- SIMILARITY: BELONGS TO FAMILY 35 OF GLYCOSYL HYDROLASES.
DR EMBL; AJ012796; CA010173.1;
DR InterPro: IPR000922; Gal_Lectin.
DR InterPro: IPR001944; Glyco_hydro_35.
DR Pfam: PF02140; Gal_Lectin; 1.
DR Pfam: PF01301; Glyco_hydro_35; 1.
DR PRINTS: PR00742; GLHYDRLASE35.
DR PRODOM: PD005612; Gal_Lectin; 1.
DR PROSITE: PS01182; GLYCOSYL_HYDROL_F35; 1.
DR PROSITE: PS50228; SUEL_LECTIN; 1.
KW Glycosidase; Hydrolase; Signal.
FT SIGNAL 1 25 POTENTIAL.
FT CHAIN 26 838 SS-GALACTOSIDASE.
SQ SEQUENCE 838 AA; 92924 MW; AC2E11ABFA417762 CRC64;
```

Query Match 74.98; Score 2941; DB 10; Length 838;
Best Local Similarity 72.08; Pred. No. 3.2e-190;
Matches 520; Conservative 97; Mismatches 101; Indels 4; Gaps 2;

```
QY 1 MLRTNVLVLLVICLLDFSSVKASVYDDRAIIINGKRKILISGSIHYPRSTPOMWPDLI 60
DB 6 ILMLNVLVLL---LGSWVFGTSASVYDHRALIVNGORRLISGSIHYPRSTPEWPFGLI 62
QY 61 QKAKGGDLVETIYVFWNGHEPSPGKYNFEGRYDLVRFIKMVRAGLYVNLRIQGYVCAE 120
DB 63 QKAKEGGVDTIQTIVFWNGHEPQOGKYFEGRYDLVRFIKLVHQAGLYVHLRVGPYCAE 122
QY 121 WNFEGFPVWLKYYVPGMEFRINQPFKVMQGFVOKIYNNMKSENLFESQGGPIIQAQIEN 180
DB 123 WNFEGFPVWLKYYVPGISFRDINGPFKAAMQKFTAKIYNNMKKAERLYETQGGPIIILSQEN 182
QY 181 EYGVVEVEIGAPGKAYTKAAQMAVGLTKTGVPTMCKOEADPDVIDTCNGFCYCEGFRPN 240
DB 183 EYGVMEVELGAPGKSYAQAAKMAVGLDTGVPTMCKQDDAPDPIINACNGFYCDYFSPN 242
QY 241 KPYKPKMWTETVGTWYTKFGGPIPORPAEDIAFVARFVQNGSFFNYMYHGTNFGRT 300
DB 243 KAYKPKIWTETAWTAFWTFGFGNPVPRPAEDLAFSAKFIQKGGSFINYMYHGTNFGRT 302
QY 301 SSGLFATSYDYDAPDEYGLLNEPKYCHLRLHKAIKLSEPALVSSYAAVTSLSGNSQEA 360
DB 303 AGGFATSYDYDAPDEYGLLRQPKWHLKDLHRAIKLCEPALVSGDPVATLVHQHQA 362
QY 361 HVYRSKGACAALNSYDSRVSVKVTQNRPNLPWPSISILPDKCTAVNTAQNVSQSS 420
DB 363 HVFBSKAGSACAALFYNDQHSFATVSEANRHYNLPWPSISILPDKCTVNTAIGQA 422
QY 421 SIKMTAGGGLSWQSYNEETPTADSDTLTANGLWEOKNVTROSSDYLVWNTVNTVNIASNE 480
DB 423 QMKMTPTVSRGLPQWSFNEETSSYEDS--SFTVVGLEQINTTRDVS DYLVWNTVNTVNIASNE 481
QY 481 GFLKNGKDPYLTYNMAGHVLHVFNGLSGTVYGLTNDPKLTYSGNVKLRAGINKILLS 540
DB 482 KFLRGKWPWLTITMSAGHALHVFNGLQAGTAGLSLEKPKLTFSAKVNLRAGVNKILLS 541
```



```
RC STRAIN=CV. CHANDLER; TISSUE=FRUIT;
RA Trainotti L., Spinello R., Casadoro G.;
RT "Unusual beta-galactosidases with a lectin-like domain are expressed
RL in strawberry."
DR Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
EMBL: AJ278703; CAC44500.1; -.
KW Hydrolase; Glycosidase.
SQ SEQUENCE 843 AA; 93505 MW; C02EFF2C93EDBA4A CRC64;

Query Match 74.3%; Score 2915; DB 10; Length 843;
Best Local Similarity 71.8%; Pred. No. 1.8e-188;
Matches 519; Conservative 95; Mismatches 101; Indels 8; Gaps 4;

Qy 5 NVLL~--LLVICLLDFFSSVKASVSDRAIIINGKRKILISGSIHYPRSTPQMPDLIQ 61
Db 9 NVVAAALVLC--SCFASVRASVSDSKAIVINGQRRILISGSIHYPRSTPQMPDLIQ 66

Qy 62 KAKDGLDVIETVFWNGHEPSGKYNFEGRYDLVRFIKMVORAGLYVNLRIQPVCAEW 121
Db 67 RAKDGLDVIQTVFWNGHEPSGKYNFEDNDYLVKFIKLVQOAGLYVHLRIQPVCAEW 126

Qy 122 NFGGPPVWLKYYPGMEFRNNQPFKVMQGFVQKIVNMKSENLEFSOGGPIIQAQIENE 181
Db 127 NFGGPPVWLKYYPGMEFRNNQPFKVMQGFVQKIVNMKSENLEFSOGGPIIQAQIENE 186

Qy 182 YGPVWEIGAPGKAYTKWAAQMAVGLKTGVPWIMCKQEDAPDPVIDTCNGFYCEGFRPNK 241
Db 187 YGPVWEIGAPGKAYTKWAAQMAVGLKTGVPWIMCKQEDAPDPVIDTCNGFYCEGFRPNK 246

Qy 242 PYKPMWTEWGTWTKFGGPIQORPAEDIAFSAVFARVONNGSFNYHYHGGTNFGRTS 301
Db 247 AYKPMWTEWGTWTEFGGPIQORPAEDIAFSAVFARVONNGSFNYHYHGGTNFGRTS 306

Qy 302 SGLFATSDYDAPLDEYGLLNEPKYGHRLDHLKAIKLSEPALVSSAAVTSILGSLNQEAH 361
Db 307 GGFATSDYDAPLDEYGLLNEPKYGHRLDHLKAIKLSEPALVSSAAVTSILGSLNQEAH 366

Qy 362 VYRSKGACAAFLSNYSRYSVKVTQNPYLPWISILPDCIKTAVYNTAQVNSQSS 421
Db 367 VFKNSGACAAFLSNYSRYSVKVTQNPYLPWISILPDCIKTAVYNTAQVNSQSS 426

Qy 422 IKM--TPAGGLSWOSYNEETPTADSDTLTANGLEQKNVTRDSSDYLWYNTVNIAS 479
Db 427 MKMPRPVTHGGFSWQAYNDETATYSDTFTTA--GLLEQINITRDATDYLWYNTVNIAS 485

Qy 480 EGFNGKDPYLTVMAGHLVHFVNGKLSGTVYGTLDNPKLTSYGNVYKLRAGINKISLL 539
Db 486 EDFLRSGNYPVLTVMAGHLVHFVNGKLSGTVYGTLDNPKLTSYGNVYKLRAGINKISLL 545

Qy 540 SVSGLPNVGVHYDWNAGVLPVTLISGLNEGRNLAKOKWYSYKGLKGESLSLSLGS 599
Db 546 STAVGLPNVGVHYDWNAGVLPVTLISGLNEGRNLAKOKWYSYKGLKGESLSLSLGS 605

Qy 600 SVSEWVRLMAOKPLTWYKATFNAGNDPLALDMSMGKQGTWINGEGVGRHWPGYI 659
Db 606 SVSEWVRLMAOKPLTWYKATFNAGNDPLALDMSMGKQGTWINGEGVGRHWPGYI 665

Qy 660 AGDCSKCYACTEKKCOTNCQOPORWYHVPKSLKPSGNLLVVFEEGNGTGLSL 719
Db 666 AGTCECNYAGTFFSEKKCLSCNCGEASQWYHVPKSLKPSGNLLVVFEEGNGTGLSL 725

Qy 720 VRR 722
Db 726 VRR 728

RESULT 12
Q92P17 PRELIMINARY; PRT; 730 AA.
AC Q92P17;
DT 01-MAY-1999 (Tremblrel. 10, Created)
DT 01-MAY-1999 (Tremblrel. 10, Last sequence update)
```

```
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE BETA-GALACTOSIDASE PRECURSOR (EC 3.2.1.23) (LACTASE).
OS Lupinus angustifolius (Narrow-leaved blue lupine).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
EMBL: AJ278703; CAC44500.1; -.
KW Hydrolase; Glycosidase.
SQ SEQUENCE 843 AA; 93505 MW; C02EFF2C93EDBA4A CRC64;

Query Match 74.1%; Score 2906; DB 10; Length 730;
Best Local Similarity 71.6%; Pred. No. 6.1e-188;
Matches 521; Conservative 94; Mismatches 99; Indels 14; Gaps 5;

Qy 1 MLRTN---VLLLV--ICLLDFFSSVKASVSDRAIIINGKRKILISGSIHYPRSTPQ 55
Db 12 MSRRNFHVVLLLVFWVY-----VTASVYDHAIMINGORRILISGSIHYPRSTPQ 65

Qy 56 WPDLIQAKDGLDVIETVFWNGHEPSGKYNFEGRYDLVRFIKMVORAGLYVNLRIQ 115
Db 66 WPDLIQAKDGLDVIETVFWNGHEPSGKYNFEDNDYLVKFIKLVQOAGLYVHLRIQ 125

Qy 116 YCAEWNEFGGPPVWLKYYPGMEFRNNQPFKVMQGFVQKIVNMKSENLEFSOGGPII 175
Db 126 YCAEWNEFGGPPVWLKYYPGMEFRNNQPFKVMQGFVQKIVNMKSENLEFSOGGPII 185

Qy 176 AQIENEYGPVWEIGAPGKAYTKWAAQMAVGLKTGVPWIMCKQEDAPDPVIDTCNGFYCE 235
Db 186 AQIENEYGPVWEIGAPGKAYTKWAAQMAVGLKTGVPWIMCKQEDAPDPVIDTCNGFYCE 245

Qy 236 GFRPNKPYKPMWTEWGTWTKFGGPIQORPAEDIAFSAVFARVONNGSFNYHYHGGT 295
Db 246 GFRPNKPYKPMWTEWGTWTKFGGPIQORPAEDIAFSAVFARVONNGSFNYHYHGGT 305

Qy 296 NFGRTSSGLFATSDYDAPLDEYGLLNEPKYGHRLDHLKAIKLSEPALVSSAAVTSILG 355
Db 306 NFGRTSSGLFATSDYDAPLDEYGLLNEPKYGHRLDHLKAIKLSEPALVSSAAVTSILG 365

Qy 356 SNOEAHVYRSKGACAAFLSNYSRYSVKVTQNPYLPWISILPDCIKTAVYNTAQV 415
Db 366 SNOEAHVYRSKGACAAFLSNYSRYSVKVTQNPYLPWISILPDCIKTAVYNTAQV 424

Qy 416 NSQSSIKMTAGGLSWOSYNEETPTADSDTLTANGLEQKNVTRDSSDYLWYNTVNI 475
Db 425 NSQSSIKMTAGGLSWOSYNEETPTADSDTLTANGLEQKNVTRDSSDYLWYNTVNI 484

Qy 476 TASNEGFLLKNGKDPYLTVMAGHLVHFVNGKLSGTVYGTLDNPKLTSYGNVYKLRAGINK 535
Db 485 TASNEGFLLKNGKDPYLTVMAGHLVHFVNGKLSGTVYGTLDNPKLTSYGNVYKLRAGINK 542

Qy 536 ISLLSVSVGLPNVGVHYDWNAGVLPVTLISGLNEGRNLAKOKWYSYKGLKGESLSLHS 595
Db 543 ISLLSVSVGLPNVGVHYDWNAGVLPVTLISGLNEGRNLAKOKWYSYKGLKGESLSLHS 602
```



```
Qy 596 LSGSSSVFVWVRSMAQKOPLTWYKATNPAGGNDPLALDMSMGKGQIWIINGEYGRHW 655
    :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 603 EAGNSVFWOGSLVAKKQPLAWTKTTESAPAGNDPLALDLSMGKGVEVWVNGQSIGRHW 662
    :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Qy 656 PCYTAQGDSCSKSVAGTNEKKCOTNCQOPSORWYHVPRLKPSGNLLVVFEEWGGNPT 715
    :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 663 PGNKARGCNCNAGVTYDTKCLANCQOPSORWYHVPRLWLRSGNLLVLEEWGGDPN 722
    :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Qy 716 GISLVRRS 723
    :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 723 GIALVERT 730
    :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
RESULT 13
Q93X56 PRELIMINARY; PRT; 722 AA.
ID AC Q93X56 PRELIMINARY; PRT; 722 AA.
AC Q93X56;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE BETA-GALACTOSIDASE (EC 3.2.1.23).
GN BETA-GAL3
OS Fragaria ananassa (Strawberry).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids I; Rosales; Rosaceae; Rosoideae; Fragaria.
OX NCBI_TaxID=3747;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. CHANDLER; TISSUE=FRUIT;
RA Trainotti L., Spinello R., Casadoro G.;
RT "Unusual beta-galactosidases with a lectin-like domain are expressed
    in strawberry.";
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ278705; CAC44502.1; -
KW Hydrolase; Glycosidase.
SQ SEQUENCE 722 AA; 81158 MW; DC08461E3F380123 CRC64;

Query Match 73.9%; Score 2901.5; DB 10; Length 722;
Best Local Similarity 71.6%; Pred. No. 1.2e-187;
Matches 513; Conservative 100; Mismatches 100; Indels 3; Gaps 3;

Qy 8 LLLVICLLDFSSVKASYDDRAIIINGKRKILISGSIHYPRSTPOMWPDLIQAKDGG 67
    : : : : : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 10 MFFLLFVLSLALASVGYDHRALIVNGKRILISGSIHYPRSTPEWMPDLLQAKDGG 69
    : : : : : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 68 LDVIETVFWNGHPSGPKYFEGRYDLVRIFKVMQVAGLVNLRIGPYVCAEWNFGGFP 127
    :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 70 LDVLQTVYFWNGHPSGPKYFEDRYDLVKFKLAQQHGLYVHLRIGPYICAEMWFGGFP 129
    :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Qy 128 VMLKYVPGMEFTNNQPKVAMQGVQKIVNMKSENLFESGGPIINAOIENEGYGVPEW 187
    :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 130 VMLKYVPGIARFTDNRPFMAWEKFTQKIVYMKAEKRLUFGGGPIILSQIENEGYGVPEW 189
    :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Qy 188 EIGAPGKAYTKWAQMAVGLKTGPVIMCKQEDAPDPVIDTCNGFYCEGFRPNKPKPKM 247
    :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 190 EIGAPGKSYTQWAKMAVGLNTGVPVWCKQEDAPDPIDTCNGFYCFENFTPNKNYKPKM 249
    :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Qy 248 WTEWWTGWTYFGGPIQPPADIEDAFSVARFVQNNGSFNFYMYHGGTFNFGRTSSGLPIA 307
    :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 250 WTEIWTGWTYFEGGAVPRPAQDLAFSVAREFTONGSGSFANYMYHGGTFNFGRTAGGPPIA 309
    :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Qy 308 TSYDDYDAPLDEYGLLNEPKYGHRLDLHKAIKLSEPALVSSVAAYTSLGSNOEAHVYRSKS 367
    :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 310 TSYDIDAPLDEYGLPREPKYSHLKYTMHRAIKMAEPALLATDAASKLGNNOEAHVYQSKS 369
    :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Qy 368 GACAAFLSNYSRYSKVYTFQNRPNYLPWPWSISILPDKCTAVYNTAQVNSSSSIKMTPTA 427
    :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 370 G-CAAFLANYDTKYVPVRYTFWNKYNLPPWSISILPDKCTEVTNTARV-GQSPPTKMTPTV 427
    :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Qy 428 GGGLSWQSYNETPTADSDTLTANGLWEQKNVTRDSSDYLWYMTNVNIASNEGLKNGK 487
    :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
```

```
Db 428 -AHLSSQAYIEDVATSADDNAFTSVGLREQISLTWDNTDYLWYMTDITIGPNEQFLRTGK 486
Qy 488 DPYLTVMASAGHLVHFVANGKLSGTWYGLDNPKLTYSGNVKLRAGINKISLLSVSVGUPN 547
    :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 487 YPTLKVDASAGHALHVFINGQLSGSAYGTLAPKLFNGVGLRAGINKALLSVSVGLAN 546
    :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Qy 548 VGVHYDTWAGVGLPVTLSGLNEGRNLAOKWYKVGKLGESLSHLSSGSSSVSEWVRG 607
    :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 547 VGLHFETWNTGVLGPVTLVAGVNSGTWDMTRWQYTKIGMRGEDMSLHTVSGSSSVSEWVQG 606
    :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Qy 608 SLMAQKQPLTWYKATFNAPGGNDPLALDMSMGKGQIWIINGEYGRHWPGYIAQDCSKC 667
    :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 607 SLLAQYRLTWYKATLNPAGGNDPLALDMSMGKGQIWIINGEYGRHWPAKHAHSCGAC 666
    :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Qy 668 SYAGTFNEKKCOTNCQOPSORWYHVPRLKPSGNLLVVFEEWGGNPTGISLVRRS 723
    :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 667 YYAGTYTENKCRTCNGQOPSORWYHVPRLKPSGNLLVVFEEWGGDPFKISLVAR 722
    :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
RESULT 14
Q9SCV0 PRELIMINARY; PRT; 728 AA.
ID AC Q9SCV0 PRELIMINARY; PRT; 728 AA.
AC Q9SCV0;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE BETA-GALACTOSIDASE PRECURSOR (EC 3.2.1.23) (LACTASE).
GN BGAL12
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Gy I., Kreis M., Lecharny A.;
RT "The beta-galactosidases are encoding by a multigene family in
    Arabidopsis thaliana.";
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
CC -1- CATALYTIC ACTIVITY: HYDROLYSIS OF TERMINAL, NON-REDUCING BETA-D-
    GALACTOSE RESIDUES IN BETA-D-GALACTOSIDES.
CC -1- SIMILARITY: BELONGS TO FAMILY 35 OF GLYCOSYL HYDROLASES.
DR EMBL; AJ270308; CAB64748.1; -
DR InterPro; IPR001944; Glyco_hydro_35.
DR Pfam; PF01301; Glyco_hydro_35; 1.
DR PRINTS; PR00742; GLHYDRASE35.
DR PROSITE; PS01182; GLYCOSYL_HYDROL_F35; 1.
KW Glycosidase; Hydrolase; Signal.
FT SIGNAL 1 23 POTENTIAL.
FT CHAIN 24 728 PUTATIVE BETA-GALACTOSIDASE.
SQ SEQUENCE 728 AA; 81532 MW; 75EB099914A7BA65 CRC64;
```

```
Query Match 72.6%; Score 2848.5; DB 10; Length 728;
Best Local Similarity 69.8%; Pred. No. 4.6e-184;
Matches 502; Conservative 104; Mismatches 108; Indels 5; Gaps 4;

Qy 7 LLLVICLLDFSSVKASYDDRAIIINGKRKILISGSIHYPRSTPOMWPDLIQAKDGG 66
    :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 11 LLLGLLCCSLTCSVKAKIVTYDRAVIINGQRILLSGSIHYPRSTPEWMPDLLQAKDGG 70
    :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Qy 67 GLDVIETVFWNGHPSGPKYFEGRYDLVRIFKVMQVAGLVNLRIGPYVCAEWNFGGFP 126
    :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 71 GLDVIQTVFWNGHPSGPKYFEDRYDLVKFKIVVQQAGLVVHLRIGPYVCAEWNFGGFP 130
    :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Qy 127 PWLKYVPGMEFTNNQPKVAMQGVQKIVNMKSENLFESGGPIINAOIENEGYGVPE 186
    :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 131 PWLKYVPGVFTDNEPEKAAQKFTKIVRMKEELFTETGGPIILSQIENEGYGPTE 190
    :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Qy 187 WEIGAPGKAYTKWAQMAVGLKTGPVIMCKQEDAPDPVIDTCNGFYCEGFRPNKPKPK 246
    :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 191 WEIGAPGKAYTKWAQMAVGLKTGPVIMCKQEDAPDPVIDTCNGFYCEGFRPNKPKPK 250
    :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
```

```
QY 247 MTEVMTGWYTKFGGPIQORPAEDIAFSAVFQVNGSGFFNYMYHGGNFQRTSSGLFI 306
Db 251 MTTNNWTFTEFGGAVPYRPAEDIALSARFETQNGSGFFNYMYHGGTFNFDRT-AGEFI 309
QY 307 ATSYDYDAPLDEYGLNEPKYGHLDLHKAIKLSALYSSAAVTSLSNQEAHYRSK 366
Db 310 ATSYDYDAPLDEYGLPREPKYSHLKLHVKIKLCEPALVSADPTVTSGLGKQEAHYRFSK 369
QY 367 SGCAAFSLSNYSRYSVKVTFQNRPNLPPWSISILPCKTAVNTAOVNSOSSSI--KM 424
Db 370 S-SCAAFLSNYNTSSAARVLFGGSTDYDLPWWSVILPCKTEYNTAKVQVTSIHHMKM 428
QY 425 TPAGGLSQWSYNEETPTADSDTLTANGLWEQKNVTRDSDYLWYMTNVNIASNEGLK 484
Db 429 VPTNTPFSWGSYNEEIPISANDNGTFSQDGLVQISITRDKTDYFWLTDITISPDKEFL- 487
QY 485 NGKDPYLTVMASAGHVLHVFVNGKLSGTVYGLDNPALPKLTYSGNVKLKLAGINKISLSVSG 544
Db 488 TGEDPLLTGSGAGHALHVFVNGSLAGTAGYAGSLKPKLTFQKIKLHAGVYKIALLSAAG 547
QY 545 LPNVGVHYDTWAGVYGLPVTLSGLNEGSRLAKQKWSYKVGKLSLHSLSGSSSVV 604
Db 548 LPNVGVHYETWNTGVLGPTLVNGVNSGTWDMTKKWSYKIGTKGEALS VHTLAGSSTVEM 607
QY 605 VRGSLMAQKQPLTWYKATFNAGGNDPLALDMASMGKGQIWIINGEGVGRHWPYIAQGD 664
Db 608 KEGSLVAKKQPLTWYKSTFDSPTGNEPLALDMNTMGKQMWINGQNGIRHWPAYTARGK 667
QY 665 SKCSYAGTNEKKQCNPCGQSORWYHVPWSLKPNSGLLVFEEWGGNPTGISLVRS 723
Db 668 ERCSTAGTTEKKCLSNCGEASQWYHVPWSLKPNTNLVLEEWGGEPNGISLVKRT 726

RESULT 15
Q9SZ15 PRELIMINARY; PRT; 729 AA.
AC Q9SZ15;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DE BETA-GALACTOSIDASE (EC 3.2.1.23) (LACTASE).
GN F20818.250 OR AF4G26140.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Bevan M., Rose M., Hempel S., Entian K.-D., Hoheisel J., Mewes H.W.,
RA Mayer K.F.X., Schueller C.;
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Rose M., Hempel S., Entian K.-D., Mewes H.W., Lemcke K., Mayer K.F.X.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
CC -1- CATALYTIC ACTIVITY: HYDROLYSIS OF TERMINAL, NON-REDUCING BETA-D-
CC GALACTOSE RESIDUES IN BETA-D-GALACTOSIDES.
CC -1- SIMILARITY: BELONGS TO FAMILY 35 OF GLYCOSYL HYDROLASES.
DR EMBL; AL049483; CAB39679.1; -
DR EMBL; AL161564; CAB79469.1; -
DR InterPro; IPR001944; Glyco_hydro.35.
DR Pfam; PF01301; Glyco_hydro.35; 1.
DR PRINTS; PR00742; GLHYDRLASE3.
DR PROSITE; PS01182; GLYCOSYL_HYDROL_F35; 1.
```

KW Glycosidase; Hydrolase.
SQ SEQUENCE 729 AA; 81660 MW; F5B75526720F696D CRC64;

Query Match 72.3%; Score 2838; DB 10; Length 729;
Best Local Similarity 69.7%; Pred. No. 2.4e-183;
Matches 502; Conservative 104; Mismatches 108; Indels 6; Gaps 5;

```
QY 7 LLLLVICLLDFFSSYKASVSYDDRAILLNGKRLKILSGSIHYPRSTPOMWPDLOKAKDG 66
Db 11 ILGLCCSLSCSKVAIVTDKAVIINGQRILLSGSIHYPRSTPOMWPDLOKAKDG 70
QY 67 GLDVIETVYFWNGHEPSPGKYNFEGRYDLVRFIKMVORAGLYVNLIRIGYVCAENFNGF 126
Db 71 GLDVIQTVYFWNGHEPSPGQYVYFEDRYDLVKFKVQVQAGLYVHLRIGYVCAENFNGF 130
QY 127 PWLVKIVVPGMEFRTNNQPFKVAQGFVQKIVNMKSENLFESQGGPIIMAOIENEGPVE 186
Db 131 PWLVKIVVPGMEFRTNEPFKAAQKFEKIVRMKEEKLFTQGGPIILSQIENEGPTE 190
QY 187 WEIGAPGKAYTKWAAMAVGLKTGVPTMCKOEDAPDPVIDTCNGFYCEGPRPNKPYKPK 246
Db 191 WEIGAPGKAYTKWAAMAVGLKTGVPTMCKOEDAPDPVIDTCNGFYCEGPRPNKPYKPK 250
QY 247 MTEVMTGWYTKFGGPIQORPAEDIAFSAVFQVNGSGFFNYMYHGGTFNFDRT-AGEFI 309
Db 251 MTTNNWTFTEFGGAVPYRPAEDIALSARFETQNGSGFFNYMYHGGTFNFDRT-AGEFI 309
QY 307 ATSYDYDAPLDEYGLNEPKYGHLDLHKAIKLSALYSSAAVTSLSNQEAHYRSK 366
Db 310 ATSYDYDAPLDEYGLPREPKYSHLKLHVKIKLCEPALVSADPTVTSGLGKQEAHYRFSK 369
QY 367 SGCAAFSLSNYSRYSVKVTFQNRPNLPPWSISILPCKTAVNTAOVNSOSSSI--KM 424
Db 370 S-SCAAFLSNYNTSSAARVLFGGSTDYDLPWWSVILPCKTEYNTAKVQVTSIHHMKM 428
QY 425 TPAGGLSQWSYNEETPTADSDTLTANGLWEQKNVTRDSDYLWYMTNVNIASNEGLK 484
Db 429 VPTNTPFSWGSYNEEIPISANDNGTFSQDGLVQISITRDKTDYFWLTDITISPDKEFL- 487
QY 485 NGKDPYLTVMASAGHVLHVFVNGKLSGTVYGLDNPALPKLTYSGNVKLKLAGINKISLSVSG 544
Db 488 TGEDPLLTGSGAGHALHVFVNGSLAGTAGYAGSLKPKLTFQKIKLHAGVYKIALLSAAG 547
QY 545 LPNVGVHYDTWAGVYGLPVTLSGLNEGSRLAKQKWSYKVGKLSLHSLSGSSSVV 603
Db 548 LPNVGVHYETWNTGVLGPTLVNGVNSGTWDMTKKWSYKIGTKGEALS VHTLAGSSTV 607
QY 604 WVRGSLMAQKQPLTWYKATFNAGGNDPLALDMASMGKGQIWIINGEGVGRHWPYIAQGD 663
Db 608 KEGSLVAKKQPLTWYKSTFDSPTGNEPLALDMNTMGKQMWINGQNGIRHWPAYTARGK 667
QY 664 CSKCSYAGTNEKKQCNPCGQSORWYHVPWSLKPNSGLLVFEEWGGNPTGISLVRS 723
Db 668 ERCSTAGTTEKKCLSNCGEASQWYHVPWSLKPNTNLVLEEWGGEPNGISLVKRT 727
```

Search completed: June 24, 2002, 20:42:04
Job time: 264 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 25, 2002, 02:28:45 ; Search time 3524.51 Seconds
(without alignments)
9780.430 Million cell updates/sec

Title: US-09-701-868-4
Perfect score: 2554
Sequence: 1 aaaaaagttcaattttt.....aaaaaaaaaaaaaaaaaaaaa 2554

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

```

1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_estl:*
10: gb_est2:*
11: gb_hic:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	754.6	29.5	793	10	BI422837
2	754.2	29.5	778	10	BI423180
3	747.8	29.3	762	10	BM413054
4	665.6	26.1	704	10	BI432727
5	645.8	25.3	649	10	BF113419
6	645.6	25.3	653	9	AW030222
7	618.8	24.2	622	9	AW096565
8	613.6	24.0	620	10	BM410481
9	586.4	23.0	596	9	AW223645
10	581	22.7	589	9	AW092855
11	556.6	21.8	563	9	AW032987
12	549.4	21.5	551	10	BE431542
13	537.8	21.1	549	9	AW033711
14	534.4	20.9	536	9	AW035517
15	532.4	20.8	611	9	AI894883
16	524.4	20.5	526	10	BE436849
17	512.8	20.1	533	9	AW616819

18	503.8	19.7	528	9	AW033293	AW033293	EST276864
19	497	19.5	497	9	AW029823	AW029823	EST273078
20	490.4	19.2	494	9	AW217169	AW217169	EST295883
21	474.8	18.6	478	10	BE434106	BE434106	EST405184
22	473.8	18.6	612	9	AW030607	AW030607	EST273862
23	472.8	18.5	612	9	AW030428	AW030428	EST273683
24	452.4	17.7	455	9	AW441477	AW441477	EST310873
25	450.2	17.6	612	10	BF187952	BF187952	EST444239
26	435.6	17.1	611	10	BG097633	BG097633	EST462152
27	423.4	16.6	661	10	BF641354	BF641354	NF066E11
28	421	16.5	647	10	BI922495	BI922495	EST542399
29	414.6	16.2	610	9	AI894508	AI894508	EST263951
30	406	15.9	481	9	AW617501	AW617501	EST323912
31	404	15.8	738	10	BG841036	BG841036	MEST14-H0
32	403.8	15.8	659	9	AW696211	AW696211	NF103F08S
33	398.6	15.6	405	9	AI894534	AI894534	EST263977
34	393	15.4	789	10	BG123351	BG123351	EST468997
35	378.8	14.8	728	10	BI310794	BI310794	EST531254
36	377.8	14.8	769	10	BM412213	BM412213	EST586540
37	373.4	14.6	535	9	AB075003	AB075003	AB075003
38	370.4	14.5	674	9	AI778902	AI778902	EST259781
39	367.8	14.4	379	9	AW029933	AW029933	EST273188
40	365.4	14.3	726	10	BE660167	BE660167	O-D10 Gna
41	365.4	14.3	800	10	BG645654	BG645654	EST507273
42	364.4	14.3	604	10	BF632221	BF632221	NF029H10D
43	360	14.1	766	10	BI176439	BI176439	EST517384
44	358.4	14.0	666	9	AI055468	AI055468	Coau0004B
45	358	14.0	358	9	AW933150	AW933150	EST358993

ALIGNMENTS

RESULT	1
BI422837	
LOCUS	BI422837 793 bp mRNA linear EST 16-AUG-2001
DEFINITION	EST533503 tomato callus, TAMU Lycopersicon esculentum cDNA clone
ACCESSION	BI422837
VERSION	BI422837.1 GI:15197578
KEYWORDS	EST.
SOURCE	tomato.
ORGANISM	Lycopersicon esculentum
REFERENCE	1 (bases 1 to 793)
AUTHORS	Alcala,J., Vrebalov,J., White,R., Matern,A.L., Vision,T., Holt,I.E., Lilang,F., Upton,J., Craven,M.B., Bowman,C.L., Ahn,S., Ronning,C.M., Fraser,C.M., Martin,G.B., Tanksley,S.D. and Giovannoni,J.
TITLE	Generation of ESTs from tomato callus tissue
JOURNAL	Unpublished (1999)
COMMENT	Contact: CUGI Clemson University Genomics Institute Clemson University 100 Jordan Hall, Clemson, SC 29634, USA Email: http://www.genome.clemson.edu/orders/index.html . Location/Qualifiers 1. .793 /organism="Lycopersicon esculentum" /cultivar="TA496" /db_xref="taxon:4081" /clone="cLEC71F18" /clone_lib="tomato callus, TAMU" /tissue_type="callus" /dev_stage="25-40 days old" /lab_host="XL1-Blue MRF" /note="Vector: pBlueScript SK(-); Site_1: EcoRI; Site_2: of seedlings 7-10 days post-germination were excised, cut at both ends and placed on MS medium with no selection. Mixed callus was harvested at 25 and 40 days and included
FEATURES	source

BASE COUNT	238 a	159 c	177 g	219 t	undifferentiated masses, Tomato Callus EST Library"
ORIGIN					
Query Match	29.5%	Score 754.6;	DB 10;	Length 793;	
Best Local Similarity	98.1%	Pred. No. 6.7e-96;			
Matches 774;	Conservative 0;	Mismatches 14;	Indels 1;	Gaps 1;	
Qy	1051	ttgagagacttacataaagctatcaagctatctgaacggctttagtttcatctatgatct	1110		
Db	5	TTGAGAGACTTTACATAAAGCTATCAAGCTATCTGAACCGGCTTTAGTTTCATCATATGCT	64		
Qy	1111	gcgctgactgctcttgaagtaatacaagagctcatgtttataagatcaaaatctcgagct	1170		
Db	65	CGGTGACTAGCTTTGGAAGTAATCAAGAGGCTCATGTTATAGATCAAAATCTGGAGCT	124		
Qy	1171	tgtctgcttttttatccaaactatgactctctagatattcagtaaaagtcacaccttcagaat	1230		
Db	125	TGTGCTGCTTTTTATCCAACTATGACTCTAGATATTCACTAAAAGTCACCTTTCAGAAT	184		
Qy	1231	aggccatacaatctgcctccatggtccatcagcattcttcccgactgcgaacatgcgcgtt	1290		
Db	185	AGGCCATACAATCTGCCCTCCATGTGCTCCATCAGCATCTTTCGCCAGTCGCAAACTGCCGTT	244		
Qy	1291	tacaacactgcacaggttaactctcaagctcgcagcataaagatgcagcctgcaggtggt	1350		
Db	245	TACAACACTGCACAGGTTAACTCTCAAAGCTCGAGCATAAAGATGACGCTTCGAGGTGGT	304		
Qy	1351	ggattctctggcagtcatacaaatgaagaaacgcctactgcctgctgatgcagcgatcacactt	1410		
Db	305	GGATTGCTTTGGCAGTCATACAATGAAGAAACGCCCTACTGCTGATGACAGCGATACACTT	364		
Qy	1411	acagctaacgactatgggaacagaaaaacgtcacagagattcatcagactatctgtgg	1470		
Db	365	ACAGCTAACGGACTATGGGAACAGAAAACGCTCACAGAGATTCAACAGATCTCTGTGG	424		
Qy	1471	tacatgcaaatgtaaatatagcatctaataagaagatttctaaagaacgggaagatcct	1530		
Db	425	TACATGCAAAATGTAAATATAGCATCTAATGAAGGATTTCTAAAAGACGGAAGATCCT	484		
Qy	1531	tatctactgttatccgctggctcatgctcttgatgttttcgtaatttcgtaattgaaactatca	1590		
Db	485	TATCTACTGTTATTCGCTGGTCATGCTCTGCATGTTTTCGTCATATGGAATACTATCA	544		
Qy	1591	ggaactgttatggtacattggtataaataccaaacttcacatacagtggaacgctgaagtta	1650		
Db	545	GGAACGTGTTATGTTACATTGGATAATCCAAAACCTTACATACAGTGGCAACGTTGAAGTTA	604		
Qy	1651	agaactggtattacaagattctctgcctcagtggttccgctggtctcccg-acagctgg	1709		
Db	605	AGAGCTGGTATTACAAGATTTCTCTGCTAGTGTCTTCGTTGCTCTCCGAAAACCTGG	664		
Qy	1710	cgtcattatgatcatggaatgcagagttcttagtccagtcagctgagcgctctcaa	1769		
Db	665	CGTGCATTATGATACATGGAATGCGAGAGTTCTTAGGTCAGTCAGTTGAGCGGGTCTCA	724		
Qy	1770	tgaagggttcaagaacttggcgaacagaaatggtcttacaaggttggtctgaaagcgga	1829		
Db	725	TGAAGGGGTCAAGAACTTGGCGAAACAGAAATGCTTCTTACAAGGTGCTGTGTAAGGCGA	784		
Qy	1830	atcgtttaag 1838			
Db	785	ATCTGTTAAG 793			
RESULT 2					
BI423180					
LOCUS	BI423180	778 bp	mRNA	linear	EST 16-AUG-2001
DEFINITION	EST533846 tomato callus, TAMU Lycopodium esculentum cDNA clone				
ACCESSION	CU676119 5' end, mRNA sequence.				
VERSION	BI423180.1	GI:15198218			

KEYWORDS
SOURCE ORGANISM

EST.
tomato.
Lycopersicon esculentum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asterales; euasterids I; Solanales; Solanaceae; Solanum;
Lycopersicon.

REFERENCE
AUTHORS

1 (bases 1 to 778)
Alcala,J., Vrebalov,J., White,R., Matern,A.L., Vision,T., Holt,I.E.
, Liang,F., Upton,J., Craven,M.B., Bowman,C.L., Ahn,S., Ronning
,C.M., Fraser,C.M., Martin,G.B., Tanksley,S.D. and Giovannoni,J.
Generation of ESTs from tomato callus tissue
Unpublished (1999)

TITLE
JOURNAL
COMMENT

Contact: CUGI
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Email: <http://www.genome.clemson.edu/orders/index.html>.

FEATURES
source

1..778
Location/Qualifiers
/organism="Lycopersicon esculentum"
/cultivar="TA496"
/db_xref="taxon:4081"
/clone="cLEC76119"
/clone_lib="tomato callus, TAMU"
/tissue_type="callus"
/dev_stage="25-40 days old"
/lab_host="XLI-Blue MRF"
/note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
XhoI; supplier: Giovannoni laboratory; cLEC - Cotyledons
of seedlings 7-10 days post-germination were excised, cut
at both ends and placed on MS medium with no selection. Cut
Mixed callus was harvested at 25 and 40 days and included
undifferentiated masses. Tomato Callus EST Library"

242 a 144 c 199 g 193 t

BASE COUNT
ORIGIN

Query Match	29.5%;	Score	754.2;	DB	10;	Length	778;
Best Local Similarity	98.8%;	Pred. No.	7.7e-96;				
Matches	770;	Conservative	0;	Mismatches	8;	Indels	1;
Gaps							
QY	1587	atcaggaaactgtttatgggtacatttgataatccaaaacttacacacagtggaacgtgaa	1646				
Db		1 ATCAGGAAGCTGTTATGGTACATTGGATATCCAAAACCTTACATACAGTGGCAACGTGAA	60				
QY	1647	gttaagagctggtattataacaaagattctctgctcagtggtttcccttggtctccgaaacgt	1706				
Db		61 GTTAAAGAGCTGGTATTAAACAAGATTTCTCTGCTCAGTGTMTTCCGTGGTCTCCGAAACGT	120				
QY	1707	tggcgtgcattatgatacatggaatcgagaggttctaggtccaggttcagcttgagcggctct	1766				
Db		121 TGGCGTGCATTATGATACATAGGAATGAGAGGATTCTAGTCCACGTTCAGCGGTCT	180				
QY	1767	caatgaaggggtcgaagaacttggcgaaacagaaatgggtcttacaaggttggtctgaaagg	1826				
Db		181 CAATGAAGGGTCAAGAAACCTTGGCGAAAACAGAAATGGTCTTTACAAGGTTGGTCTGAAAGG	240				
QY	1827	cgaatcgttaagcttccactccttaagtgaggagttctctgttgaaatggattccaggttcc	1886				
Db		241 CGAATCGTTAAGTCTTCACCTCTTAAGTGGGAGTTCTTCTGTTGAATGGGTTTGAGGTTT	300				
QY	1887	actaaatggctcaaaagcagccctgacttgggtacaaaggtacatttaacgcgctcgagg	1946				
Db		301 ACTAGTGGCTCAAAAGCAGCCCTGACTTGGTTACAAAGGCTACATTTAACGCGCTGGAGG	360				
QY	1947	aaatgataccactagcttttagacatggccaagtattggaaaaggtcgaatatggataaatgg	2006				
Db		361 AAATGATCCACTAGCTTTAGACATGGCAAGTATGGGAAAGGTCAGATATGGATAAATGG	420				
QY	2007	tgaagcgttagctgcgcatttggccttggatacatagcaaaagcgcactgcagcaaatgcag	2066				
Db		421 TGAAGCGCTAGTTCGCAATTGSCCTTGGATACATAGCAAGCGGCACATGCAGCAAAATCGAG	480				

[illegible]

```

QY 2067 ttatgctggaacgttcaacgagagaagtcggaactaaactgagacacaccttctcagag 2126
DB 481 TTATGCTGGAACGTTCAACGAGAAGAAGTCCAGACTAACTGCGGACAACTTCTCAGAG 540
QY 2127 atggtaccattgttccacgactgctggtgaaacccaagtgaaactttagtagtattcga 2186
DB 541 ATGGTACCATTGTTCCACGATCGTGGCTGAACCAAGTGGAACTTGTAGTAGTATCGA 600
QY 2187 agaatgggagggttaattccacaggaattctctctagtcagagatcaagataaagaactcg 2246
DB 601 AGAATGGGAGGTAAATCCACAGCAATTTCTAGTCAGGAGATCAAGATAAAGAACTCG 660
QY 2247 aaaaagtaaaactgttccagtaactatggtgcttgaattcgcgcgcaaaaaatacatacaag 2306
DB 661 AAAAGTAAAACTGTTTCAGTAACATATGCTGCTTGAATTCGCGCCGCAAAAATACATACAG 720
QY 2307 aagctaacaatgaggtcagttcagtttcaaatgcaactgcaataaataacattagaagataa 2365
DB 721 AAGCTAAATGAGGCTACAGTTTCAATTCAGCTGAAT-AAACATTTAGAAGATAA 778

RESULT 3
BM413054 762 bp mRNA linear EST 22-JAN-2002
LOCUS EST387381 tomato breaker fruit Lycopersicon esculentum cDNA clone
DEFINITION CLEG62G13 5' end, mRNA sequence.
ACCESSION BM413054
VERSION BM413054.1 GI:18264684
KEYWORDS EST.
SOURCE tomato.
ORGANISM Lycopersicon esculentum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; euasterids I; Solanales; Solanaceae; Solanum;
Lycopersicon.
REFERENCE 1 (bases 1 to 762)
AUTHORS Alcalá,J., Vrebalov,J., White,R., Vision,T., Karamycheva,S.A., Tsai,
J., Bougri,O., Kirkness,E., Utterback,T., Van Aken,S., Ronning,
C.M., Fraser,C.M., Martin,G.B., Tanksley,S.D. and Giovannoni,J.
Generation of ESTs from tomato fruit tissue, breaker stage (2002)
JOURNAL Unpublished (2002)
COMMENT Contact: CGI
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Email: http://www.genome.clemson.edu/orders/index.html
This clone is available through the Clemson University Genomics
Institute
Seq primer: T3.
Location/Qualifiers
1..762
/organism="Lycopersicon esculentum"
/cultivar="TA496"
/db_xref="taxon:4081"
/clone="CLEG62G13"
/tissue_lib="tomato breaker fruit"
/tissue_type="pericarp"
/dev_stage="breaker"
/lab_host="SOLR"
/Note="Vector: pBluescriptSKmCudapt; Site_1: EcoRI;
Site_2: XhoI; supplier: Boyce Thompson Institute;
sequencing: The Institute for Genomic Research. Fruit
were harvested at the breaker stage (first sign of
lycopene accumulation on the blossom end of fruit). Fruit
were cut in half and the seeds and locules were discarded
prior to freezing the pericarp."
BASE COUNT 228 a 131 c 182 g 221 t
ORIGIN

Query Match 29.3%; Score 747.8; DB 10; Length 762;
Best Local Similarity 99.6%; Pred. No. 6e-95;

```

```

Matches 760; Conservative 0; Mismatches 2; Indels 1; Gaps 1;
QY 291 ctggaatggacatgagccttctcctggaataataatttgaaggaagatatgatcttgt 350
DB 1 CTGAATGGACATGAGCCTTCTCTCTGAAATAATAATTTGAAGGAAGATATGATCTGT 60
QY 351 tagattcacaataatggtacaaagagcaggactttatgtcaatttaccgtattggccctta 410
DB 61 TAGATTTCATCAAAATGGTACAAAGAGCAGGACTTTATGTCAATTTAGCTATTGGCCCTTA 120
QY 411 cgtctgctgtaagtgaactttggggattccctgtttggctaaataatagtgcctggtat 470
DB 121 CGTCTGTGCTGAATGAACTTTGGGGATTCCCTGTTGGCTAAATATATGTGCTGGTAT 180
QY 471 ggaatttagaacaacaatcagccttttaagtggtgctcatgcaaggatttttcagaaaaa 530
DB 181 GGAATTTAGAACAAACAATCAGCCTTTTAAGTGGCTATGCAAGGATTTGTTTCAGAAAT 240
QY 531 agtaacatgatgaagtcagaaaaatttgtttgaatctcaaggaggagccaataattatggc 590
DB 241 AGTCAACATGATGAAGTCAGAAAAATTTGTTTGAATCTCAAGGAGGACCAATAATTATGGC 300
QY 591 ccagatacaaaaatgagtatgaccagtagaataatgggaattgggtcctctgttaagctta 650
DB 301 CCAGATAGAAAAATGAGTATGGACAGTAGAATGGGAAATTTGGTCTCTCGTAAAGCTTA 360
QY 651 tacaaaatggcgagctcaaaatggctgtagtttgaaaaactggtgccatgggatcatgtg 710
DB 361 TACAAAATGGCAGCTCAAAATGGCTGTAGGTTTGAAAACCTGGTGTCCCATGGATCATGTG 420
QY 711 taagcaagaggatgctcctgatactctgattgatacttgaatggcttacttcgcgaag 770
DB 421 TAAGCAAGAGGATGCTCCTCATCTGTGATTGATGATCTGTAATGGCTTCTACTGCGAAGG 480
QY 771 gtccgtccttaataagccttacaacctaataatgtggacagaagtagtgactggctggtta 830
DB 481 GTTCCCTCTCTTAATAAGCCTTACAAACCTAAAAATGTGGACAGAAGTATGGACTGGTGTA 540
QY 831 tacgaattcgggtggtcccaattcctcaaacagcagccgacagacattgcatttcagttgc 890
DB 541 TAGCAAAATTCGGTGGTCCCAATTCCTCAAGACCAGCGGAGACATGTCATTTTCAGTTGC 600
QY 891 caggtttgttcagacaacaatgggttcattcttcaattactacatgataatggaggaacaaa 950
DB 601 CAGGTTTGTTCAGACAACAATGGTTTCATCTTCAATTTACTACATGATCATGGAGGAACAAA 660
QY 951 ttttgccggacatacatcagggttttctcattgcaactagctagattatgactcctct 1010
DB 661 TTTTGGCCGGACATCATCAGGGCTTTTCAFTTCAACTAGCTACGATTATGATGCTCCTCT 720
QY 1011 cgaatgaatggtgtgctgaatgaacaaagtatgggcaattg 1053
DB 721 CGATGAATATGGGGTCTGTAATGAACC-AAGTATGGCAGCTTG 762

RESULT 4
BI432727 704 bp mRNA linear EST 21-AUG-2001
LOCUS EST535488 P. infestans-challenged leaf Solanum tuberosum cDNA clone
DEFINITION PPCAU39 5' sequence, mRNA sequence.
ACCESSION BI432727
VERSION BI432727.1 GI:15257417
KEYWORDS EST.
SOURCE potato.
ORGANISM Solanum tuberosum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
REFERENCE 1 (bases 1 to 704)
AUTHORS Restrepo,S., Griffiths,H.M., Smart,C.D., Cho,J., Chieningo,A.,
Bougril,O., Buell,C.R., Ronning,C.M., Fry,W.E. and Baker,B.
Generation of ESTs from Potato Leaves Challenged with Phytophthora
infestans, Compatible Interaction

```



```
QY 471 ggaattagaacaaacacgctctttaaagtggtgctatgcaagattgttcagaaat 530
|||||
Db 181 GGAATTTAGAACAAACATCGCCTTTTAAAGTGGCTATGCAAGGATTGTTCACAAAT 240
|||||
QY 531 agtcaatgatgaagtcagaaattgtttgaatcacaagagagacaaataattatggc 590
|||||
Db 241 AGTCAACATGATGAAGTCAGAAATTTGTGTAATCTCAAGGAGGACCAATAATTATGCG 300
|||||
QY 591 ccagatacaaaatgatgtagcaccagtagaattggaaattgctcctcggtaaagctta 650
|||||
Db 301 CCAGATAGAAATGATGATGGACCAAGTAGAATGGAAATGGTCTCCGTTAAAGCTTA 360
|||||
QY 651 tacaataaggcgagctcaaaatggctgtaggtttgaaaactggtgtcccaatggtatcgtg 710
|||||
Db 361 TACAATAAGGGCAGCTCAAAATGGCTGTAGGTTTGAAACTGTGCTCCCATGATCATGTCG 420
|||||
QY 711 taagcaagagatgctcctgatcctgtagtatactgtagtatactgtagtatactgtagt 770
|||||
Db 421 TAAGCAAGAGGATGCTCTGATGCTGATGCTGATGATGCTGATGCTGATGCTGATGCTG 480
|||||
QY 771 gtccgctcctaagccttcaaacctcaaacctcaaacctcaaacctcaaacctcaaacctca 830
|||||
Db 481 GTTCGCTCTAATAGCCCTTCAAACTTAAATGTTGAGCAGAGATGAGCTGGCTGTTA 540
|||||
QY 831 tacgaattcgtggtgtccaaattcctcaaacctcaaacctcaaacctcaaacctcaaacctca 890
|||||
Db 541 TACGAATTCGGTGTGCTCAAAATTCCTCAAAAGCAGCCGAGACATTCGATTTTCAGTTTC 600
|||||
QY 891 caggtgtgtcagaacaaatggttcattcttcaattactatcatatcatatcatatcatatcat 939
|||||
Db 601 CAGGTTTGTTCAGAACAAATGGTTCAATCTCTCAATTAATTACTATCATATCATATCAT 649
|||||

RESULT 6
AW030222 553 bp mRNA linear EST 18-MAY-2001
LOCUS EST273477 tomato callus, TAMU Lycopersicon esculentum cDNA clone
DEFINITION cLEC19L18, mRNA sequence.
ACCESSION AW030222
VERSION AW030222.1 GI:5888978
KEYWORDS EST.
SOURCE tomato.
ORGANISM Lycopersicon esculentum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; euasterids I; Solanales; Solanaceae; Solanum;
Lycopersicon.
REFERENCE 1 (bases 1 to 553)
AUTHORS Alcala,J., Vrebalov,J., White,R., Matern,A.L., Vision,T., Holt,I.E.,
Liang,F., Upton,J., Craven,M.B., Bowman,C.L., Ahn,S., Ronning
,C.M., Fraser,C.M., Martin,G.B., Tanksley,S.D. and Giovannoni,J.
Generation of ESTs from tomato callus tissue
Unpublished (1999)
COMMENT Contact: CUGI
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Email: http://www.genome.clemson.edu/orders/index.html
5 prime sequence.
Location/Qualifiers
1..653
/organism="Lycopersicon esculentum"
/cultivar="TA496"
/db_xref="taxon:4081"
/clone="cLEC19L18"
/clone_lib="tomato callus, TAMU"
/tissue_type="callus"
/dev_stages="25-40 days old"
/lab_host="XL1-Blue MRF"
/note="vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
XhoI; supplier: Giovannoni laboratory; cLEC - Cotyledons
of seedlings 7-10 days post-germination were excised, cut
```

at both ends and placed on MS medium with no selection.
Mixed callus was harvested at 25 and 40 days and included
undifferentiated masses. Tomato Callus EST Library"

```
BASE COUNT 186 a 123 c 174 g 169 t 1 others
ORIGIN
Query Match 25.3%; Score 645.6; DB 9; Length 653;
Best Local Similarity 99.2%; Pred. No. le-80;
Matches 648; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
QY 1553 gtcagtcttgcagtgttttcgtcaaatggaactatcaggaactgtttatggtacattgg 1612
|||||
Db 1 gTCATGCTCTGTGATGCTTTTCGTCAAATGGAAACTATCAGGAAGCTGTTATGTGTACATTGG 60
|||||
QY 1613 ataatccaaactcatatcacagtggcaactgaagttaagcttaagctgtattacaagattt 1672
|||||
Db 61 ATAAATCCAAACATTTACATACAGTGGCAAGCTGAAGTTAAGAGCTGTTATTAACAAGATT 120
|||||
QY 1673 ctctgctcagtggtttccgttggtctcccgaacgttgctgcattatgatacatggaatg 1732
|||||
Db 121 CTCTGCTCAGTGTTCCTGTTGGTCTCCCGACAGTTCGCGTGCATTATGATACATGGAATG 180
|||||
QY 1733 caggagttctaggtccagtcacgttgagcgttcctcaatgaagggtcaagaacttgccga 1792
|||||
Db 181 CAGGAGTTCCTAGTGTCCAGTCAGCTTGAGCGGTCTCAATGAAGGGTCAAGAAACTTGGCGA 240
|||||
QY 1793 aacagaataatggtcttcaaaagttggtctgaagggaactcgttaactctcactcttaa 1852
|||||
Db 241 AACAGAAATGGTCTTTACAAGGTGGTCTGAAAGGGCAATCGTTAAGTCTTCATCTCTTAA 300
|||||
QY 1853 gtgggagttctctgttggaatgggttcgaggttcactaatggtcctcaaaagcagccctga 1912
|||||
Db 301 GTGGGAGTTCTTCTGTTGAATGGGTTTCAGAGTTTCACTAGTGGCTCAAAAGCAGCCCTGA 360
|||||
QY 1913 ctgtgtacaaggctacatttaacgcgcctggaggaaatgatccactagtttagacatgg 1972
|||||
Db 361 CTGTGTACAAAGCTACATTTAAACGCGCTTGAGAGAAATGATCCACTTAGCTTTAGACATGG 420
|||||
QY 1973 caagtatggaaaagtcagatatgataaatggtgaagcgtagtcgccatgcctg 2032
|||||
Db 421 CAAGTATGGAAAGCTCAGATATGGATATGTAATGGTGAAGCGGTAGGTCCCATTTGCCCTG 480
|||||
QY 2033 gatacatagcacaagcgcactgcagcaaatgcagttatgctggaacgttcaacgaaga 2092
|||||
Db 481 GATACATACACAAAGCGAGCTGCAGCAAAATGCAGTTATGCTGGAACTTCAACAGAGA 540
|||||
QY 2093 agtcggaactaactgcggacaacctctcagagatggttaccatgttcacacgcatcgcc 2152
|||||
Db 541 AGTGCCAGAGCTAACTCGGACAACTTCTCAGAGATGGTACCATTGTTCCACGATCGTGC 600
|||||
QY 2153 tgaaccacagtgaaactgttagtagtattcgaagaatggagagtaatcca 2205
|||||
Db 601 TGAACCAAGTGGAAACTTGTAGTAGTATTCGAAGAATGGGGAGGTAAATCCA 653
|||||

RESULT 7
AW096565 622 bp mRNA linear EST 18-MAY-2001
LOCUS EST289745 tomato mixed elicitor, BTI Lycopersicon esculentum cDNA
DEFINITION clone cLET38P6, mRNA sequence.
ACCESSION AW096565
VERSION AW096565.1 GI:6062160
KEYWORDS EST.
SOURCE tomato.
ORGANISM Lycopersicon esculentum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; euasterids I; Solanales; Solanaceae; Solanum;
Lycopersicon.
REFERENCE 1 (bases 1 to 622)
AUTHORS D'Ascenzo,M., He,X., Lyman,J., Holt,I.E., Liang,F., Upton,J.,
Ronning,C.M., Craven,M.B., Fujii,C.Y., Bowman,C.L., Nierman,W.,
```


Fraser,C.M., Venter,J.C., Martin,G.B., Tanksley,S.D. and Giovannoni
,J.
Generation of ESTs from tomato leaf tissue
Unpublished (1999)
Contact: CUGI
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Email: <http://www.genome.clemson.edu/orders/index.html>
5 prime sequence.

TITLE
JOURNAL
COMMENT

LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

BM410481
EST584808 tomato breaker fruit Lycopersicon esculentum cDNA clone
CLEG53G24 5' end, mRNA sequence.
BM410481
BM410481.1 GI:18262111
EST.
tomato.
Lycopersicon esculentum

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; euasterids I; Solanales; Solanaceae; Solanum;
Lycopersicon.

1 (bases 1 to 620)
Alcala,J., Vrebalov,J., White,R., Vision,T., KaramyCheva,S.A., Tsai
,J., Bougri,O., Kirkness,E., Utterback,T., Van Aken,S., Renning
,C.M., Fraser,C.M., Martin,G.B., Tanksley,S.D. and Giovannoni,J.
Generation of ESTs from tomato fruit tissue, breaker stage (2002)
Unpublished (2002)
Contact: CUGI
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Email: <http://www.genome.clemson.edu/orders/index.html>
This clone is available through the Clemson University Genomics
Institute
Seq primer: T3

Location/Qualifiers
1. .620
/organism="Lycopersicon esculentum"
/cultivar="Rio Grande PtoR"
/db_xref="taxon:4081"
/clone="cLEP38P6"
/clone_lib="tomato mixed elicitor, BTI"
/tissue_type="leaf"
/dev_stage="4-6 week old plants"
/lab_host="XLI-Blue MRF"
/note="Vector: pBlueScript SK(-); Site 1: EcoRI; Site 2:
XhoI; cLEP - Inoculated with a variety of disease response
elicitors. Plants exposed to 2,6 dichloroisonicotinic
acid, BTH, jasmonic acid, ethylene, fenthion, EIX,
okadaic acid, or systemin prior to tissue harvest. EcoRI
site was destroyed during cloning."

BASE COUNT 204 a 79 c 129 g 210 t
ORIGIN

Query Match 24.2%; Score 618.8; DB 9; Length 622;
Best Local Similarity 99.7%; Pred. No. 5.6e-77;
Matches 620; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 8 gtttcatttttttataaaataaaataaaattcatttttttgaaatgggaaaaaatgc 67
Db 1 GTTTCAATTTTTTTTCTAAATAAAAAAAATTCATTTTTTTTGAATGGAAAAATGC 60
QY 68 taaggactaagtgtgtgtattagttattgtttattgattttttttctcagtgga 127
Db 61 TAAGGACTAATGTGTGTGTATTAGTATTGTGTATTGTGTATTGTGTATTGTGT 120
QY 128 aagctagtgtttcttatgatgcagagctataatcatataaatgggaaaaaaattcttta 187
Db 121 AAGCTAGTGTCTTATGATGACAGAGCTATATCATATAATGGAAAGAAAAATCTTA 180
QY 188 ttctgttcaattcattcaagaagcactccacagatggtgcctgattcttatcaaa 247
Db 181 TTCTGTGTTCAATTAATFCCAAAGAGCACCACAGATGTGGCTGATCTTATACAA 240
QY 248 aggcataaagatggagccttagatgtttatgaaacttatgttttctggaatggacatgagc 307
Db 241 AGGCTAAGATGGAGGCTTAGATGTATTGAACATTAATGTTTCTGGAATGGACATGAGC 300
QY 308 ctctcctggaaaaataatttttgaaggagatgatgatctgttagattcaaaaatgg 367
Db 301 CTCTCTCTGSAATAATAATTTTGAAGGAGATATGATCTGTGTAGATCATCAAAATGG 360
QY 368 tacaagagcagactttagtcaatttcacgtattgtggcccttacgtctgtgctgaatgga 427
Db 361 TACAAAGACGAGGACCTTATGTCAATTTACGTATTGGCCCTTACGTCGTGTGAAATGGA 420
QY 428 actttggggattccctgtttggctaaaataatgtgcctggtatggaattagaacaaca 487
Db 421 ACTTTGGGGATTCCTGTGTGCTAAATATGTGCTGTGTAATTTAGAACAAACA 480
QY 488 atcagccttttaagtggtgtatgcaaggatttgtttcagaaaaatagtcacatgatgaagt 547
Db 481 ATCAGCCCTTTAAGTGGGTATGCAAGGATTTGTTTCAAGAAATAGTCAACATCATGAAGT 540
QY 548 cagaaaattttttagtcatcgaaggagcaataattatggccagatatacaaaatgaat 607
Db 541 CAGAAAATTTGTTGAATCTCAAGGAGGACCAATAATATTGCCCCCAGATAGAAAATGAGT 600

QY 608 atggaccagtagaattgggaaat 629
|||||
Db 601 ATGGACCAGTAGAATGGAAAT 622

RESULT 8

BM410481
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

EST584808 tomato breaker fruit Lycopersicon esculentum cDNA clone
CLEG53G24 5' end, mRNA sequence.
BM410481
BM410481.1 GI:18262111
EST.
tomato.
Lycopersicon esculentum

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; euasterids I; Solanales; Solanaceae; Solanum;
Lycopersicon.

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Alcala,J., Vrebalov,J., White,R., Vision,T., KaramyCheva,S.A., Tsai
,J., Bougri,O., Kirkness,E., Utterback,T., Van Aken,S., Renning
,C.M., Fraser,C.M., Martin,G.B., Tanksley,S.D. and Giovannoni,J.
Generation of ESTs from tomato fruit tissue, breaker stage (2002)
Unpublished (2002)
Contact: CUGI
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Email: <http://www.genome.clemson.edu/orders/index.html>
This clone is available through the Clemson University Genomics
Institute
Seq primer: T3

FEATURES
source

1. .620
/organism="Lycopersicon esculentum"
/cultivar="TA496"
/db_xref="taxon:4081"
/clone="CLEG53G24"
/clone_lib="tomato breaker fruit"
/tissue_type="Pericarp"
/dev_stage="breaker"
/lab_host="SOLR"
/note="Vector: pBlueScriptSMCUadapt; Site_1: EcoRI;
Site_2: XhoI; supplier: Boyce Thompson Institute;
sequencing: The Institute for Genomic Research. Fruit
were harvested at the breaker stage (first sign of
lycopene accumulation on the blossom end of fruit). Fruit
were cut in half and the seeds and locules were discarded
prior to freezing the pericarp."
BASE COUNT 188 a 128 c 141 g 163 t
ORIGIN

Query Match 24.0%; Score 613.6; DB 10; Length 620;
Best Local Similarity 99.4%; Pred. No. 2.9e-76;
Matches 616; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1210 gtaaaagtcacctttcagaataggccatacaatctgcctccatggtccatcagcattctt 1269
Db 1 GTAAAAAGTCACCTTTCAGAAATAGGCCATCAATCTCCCTCCATGGTCCATCAGCATCTT 60
QY 1270 cccgactgcacaaactgccgtttacacactgcacaggttaactctcaagctcgagcata 1329
Db 61 CCGGACTGCAAAACTGCCGTTTACACACTGCACAGTAACTCTCAAGCTCGACATA 120
QY 1330 aagatgacgcctgcaggtggtgattctcttggcagtcatacaatgaagaacgcctact 1389
Db 121 AAGATGACGCTGTCAGGTGTGATTGTCTTGGCAGTATACAAATGAAGAAACGCCTACT 180
QY 1390 gctgatgacagcgatcacattacagctaacgactatggagactatggagacaaaacgtcacaaga 1449
|||||

Db 181 GCTGATGACGAGTACACTTACAGCTAACGACATATGGGAACAGAAAAACGTCACAAGA 240
 QY 1450 gattcatcagactatctgtgggtacatgacaaatgtaaatatagcatctaaatgaaggattt 1509
 Db 241 GATTTCATCAGACTATCTGGGTACATGACAAATGTAATATAGCATCTAATCAAGGATTT 300
 QY 1510 ctaagaacggaagatccttatctcactgttatgtccgctggtcattgtcttgcatgtt 1569
 Db 301 CTAAGAAACGGAAGAGATCCTTATCTCAGTGTATGTCGCTGGTGCATGCTTGCATGTT 360
 QY 1570 ttctgcaatgaaaaactatcaggaactgtttatgtatcatatgataatccaaaacttaca 1629
 Db 361 TTCGTCATGGAAGAACTATCAGAACTGTTATGTATGATCATTTGGATAATCCAAACTTACA 420
 QY 1630 tacagtggcaactgaaagttaagaactgtgatttaacaagattctctgctcagttttcc 1689
 Db 421 TACAGTGGCAAGCTGAAGTAAAGAGCTGGTATTAACAAGATTTCTCTGCTCAGTCTTCC 480
 QY 1690 gtggctccgaaacgttggcggtgcattatgatacatggaatgcaggagttcttaggtcca 1749
 Db 481 GTTGTCTCCGGAACGTTGGCGTGCATTATGATACATGGAATGCAGGAGTTCTAGTGCCA 540
 QY 1750 gtcagttgacggtctcaatgaagggtcaagaaactggcgaacacacaaatggctttac 1809
 Db 541 GTTGTCTCCGGAACGTTGGCGTGCATTATGATACATGGAATGCAGGAGTTCTAGTGCCA 600
 QY 1810 aaggttggtctgaaaggcga 1829
 Db 601 CAGGTTGCTGAAGGCGA 620

RESULT 9

LOCUS AW223645 596 bp mRNA linear EST 18-MAY-2001
 DEFINITION EST300456 tomato fruit red ripe, TAMU Lycopersicon esculentum cDNA
 clone cLEN12J23, mRNA sequence.

ACCESSION AW223645
 VERSION AW223645.1 GI:6535329

KEYWORDS EST.
 SOURCE tomato.

ORGANISM Lycopersicon esculentum

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Asteridae; euasterids I; Solanales; Solanaceae; Solanum;

REFERENCE 1 (bases 1 to 596)

Arcal, J., Vrebalov, J., White, R., Matern, A.L., Holt, I.E., Liang, F.,
 Upton, J., Hansen, T., Craven, M.B., Bowman, C.L., Ahn, S., Ronning, C.M.,
 Fraser, C.M., Martin, G.B., Tanksley, S.D. and Giovannoni, J.

Generation of ESTs from tomato fruit tissue

Unpublished (1999)

Contact: CUGI

Clemson University Genomics Institute

Clemson University

100 Jordan Hall, Clemson, SC 29634, USA

Email: <http://www.genome.clemson.edu/orders/index.html>

5 prime sequence.

FEATURES

source

Location/Qualifiers

1..596

/organism="Lycopersicon esculentum"

/cultivar="TA496"

/db_xref="taxon:4081"

/clone="cLEN12J23"

/clone.lib="tomato fruit red ripe, TAMU"

/tissue_type="pericarp"

/dev_stages="red ripe (7-20 days post-breaker)"

/note="Vector: pBluescript SK(-); Site1: EcoRI; Site2:

XhoI; supplier: Giovannoni; Fruit were tagged at the

breaker stage (first sign of lycopene accumulation on the

blossom end of the fruit) and harvested 7 days

post-breaker (fully red-ripe), 10 days post breaker, and

20 days post-breaker (over-ripe). 20 day fruit which

showed external or internal signs of pathogenesis were

discarded. Fruit were cut in half and the seeds and
 locules were discarded prior to freezing the pericarp."

BASE COUNT 169 a 113 c 157 g 157 t

ORIGIN

Query Match 23.0%; Score 586.4; DB 9; Length 596;

Best Local Similarity 99.0%; Pred. No. 1.8e-72;

Matches 590; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1513 aagaacggaagatccttatctcactgttatgtccgctggtcattgtcttgcatgttttc 1572

Db 1 AAGAACGGAAGAGATCCTTATCTCATCTGTATGTCGCTGGTCATGCTCTGTCATGTTTC 60

QY 1573 gtcaatggaactatcaggaactgtttatgtatcatggaataatccaaaacttacacac 1632

Db 61 GTCATGGAAGAACTATCAGGAAGTGTATGATGATGGAATAATCCAAACTTACATAC 120

QY 1633 agtggcaacgtgaagttaagagctggtattaaagaagattctctcagtggtttccgtt 1692

Db 121 AGTGGCAACGTGAAGTTAAGAGCTGGTATTAACAAGATTTCTCTGCTCAGTGTTCGGTT 180

QY 1693 ggtctccgaaacgttggcggtgcattatgatacatggaatgcaggagttcaggtccagtc 1752

Db 181 GGTCTCCGGAACGTTGGCGTGCATTATGATACATGGAATGCAGGAGTTCTAGGTCCAGTC 240

QY 1753 acgttgagcgtctcaatgaagggtcaagaaactggcgaacacagaatggtcttacaag 1812

Db 241 ACGTTGAGCGGTCTCANTGAAGGGTCAAGAACTTTGGGAAACACAGAAATGGTCTTACAG 300

QY 1813 gttggtctgaaaggcgaatcgttaagtcttactccttaagtggaggttcttctgtttaa 1872

Db 301 GTTGTCTGAAAGGCGAATCGTTAAGTCTTCACTCTTAAAGTGGAGTTCTTCTGTTGAA 360

QY 1873 tgggttcaggttctactaatggctcaaaagcggccctgacttggataagagctacattt 1932

Db 361 TGGGTTCGAGGTCTACTAGTGGCTCAAAAGCAGCCCTGACTTGTGTACAAAGCTACATTT 420

QY 1933 aacgcgcctggaggaatgattcactagcttttagacatggcgaagtatgggaaaggtcag 1992

Db 421 AACGGCCTGGAGGAATGATCCACTAGCTTTAGACATGGCAAGTATGGAAGAAAGTCTAG 480

QY 1993 atattgataaatgttgaaggcgttagtgccattggcctggatagatagcaagagcgac 2052

Db 481 ATATGGATAAATGGTGAAGCGTAGTGCCTGCGCATTTGGCTGGATACATAGCACAAAGCGAC 540

QY 2053 tgcacaaatgcagttatgctggaacgttcaacgagaagaagtgcggaactaactg 2108

Db 541 TGCAGCAATGCAGTTATGCTGGAAAGCTTCAACGAGGAGAGTGCAGACTAACTG 596

RESULT 10

AW092855

LOCUS

DEFINITION

EST286035 tomato mixed elicitor, BTI Lycopersicon esculentum cDNA

clone cLET22F1, mRNA sequence.

ACCESSION

AW092855

VERSION

AW092855.1

KEYWORDS

EST.

SOURCE

tomato.

ORGANISM

Lycopersicon esculentum

Eukaryota; Viridiplantae;

Streptophyta; Embryophyta;

Tracheophyta;

Spermatophyta;

Magnoliophyta;

eudicotyledons;

core eudicots;

Asteridae;

euasterids I;

Solanales;

Solanaceae;

Solanum;

REFERENCE

1 (bases 1 to 589)

D'Ascenzo, M., He, X.,

Lyman, J., Holt, I.E.,

Liang, F., Upton, J.,

Ronning, C.M.,

Craven, M.B.,

Fujii, C.Y.,

Bowman, C.L.,

Nierman, W.,

Fraser, C.M.,

Venter, J.C.,

Martin, G.B.,

Tanksley, S.D. and Giovannoni, J.

Generation of ESTs from tomato leaf tissue

Unpublished (1999)

CONTACT: CUGI

Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Email: <http://www.genome.clemson.edu/orders/index.html>
5 prime sequence.

FEATURES
Source

Location/Qualifiers
1..589
/organism="Lycopersicon esculentum"
/cultivar="Rio Grande PTO"
/db_xref="taxon:4081"
/clone="cLET22F1"
/clone_lib="tomato mixed elicitor, BTI"
/tissue_type="leaf"
/dev_stage="4-6 week old plants"
/lab_host="XLI-Blue MRF"
/note="Vector: pBlueScript SK(-); Site.1: EcoRI; Site.2:
XhoI; cLET - Inoculated with a variety of disease response
elicitors. Plants exposed to 2,6 dichloroisonicotinic
acid, BTH, jasmonic acid, ethylene, fenthion, EIX,
okadaic acid, or systemin prior to tissue harvest. EcoRI
site was destroyed during cloning."
194 a 102 c 143 g 150 t

BASE COUNT 194 a 102 c 143 g 150 t
ORIGIN

Query Match 22.7%; Score 581; DB 9; Length 589;
Best Local Similarity 99.2%; Pred. No. 1e-71;
Matches 584; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1833 gtttaagcttcactccttaagtgaggattctctgttgtaagggttcaggttcactaat 1892
Db 1 GTTAAGCTTTCACCTCTTAAGTGGGAGTCTTCTGTGTGAATGGGTTCAGGTTCACTAGT 60
QY 1893 ggtctaaagcagccctgactgtgtacaggctacatttaacgcgcgcggagaaatga 1952
Db 61 GGCCTAAAGCAGCCCTGACTGTGTACAGGCTACATTACGCGCTTGAGGAATGA 120
QY 1953 tccactagcttagacatggcaagtatggaaaaggtcagatattgataaatgggtgaagg 2012
Db 121 TCCACTAGCTTTAGACATGGCAAGTATGGGAAAGTCAGATATGGATAAATGGTGAAGG 180
QY 2013 ctatggtcccatggctggatatacagacaggcagctcagcaaatgcagttatgc 2072
Db 181 CGTAGTCCCATTTGGCTGGATACATAGCACAGCGGACTGCAGCAATGCAGTTATGC 240
QY 2073 tggaaacttcaacgagaagaagtgcggaactaactgcggacacacttccagagatgga 2132
Db 241 TGGAAAGTTCAAGAGAGAAAGTGCAGACTAACTGCGGACAACTTCTCAGAGATGGA 300
QY 2133 ccactgtccacgatcgtgctgaaacccaagtggaaacttgttagtagtatttcgaagaatg 2192
Db 301 CCATGTTCCACGATCGTGGCTGAAACCAAGTGGAACTGTGTAGTAGTATTTCGAAGAATG 360
QY 2193 gggaggttaatacacaagaattctctagtcaggatcaagaataaagaactcgaaagt 2252
Db 361 GGGAGGTAAATCCACAGGAATTTCTCTAGTCAGGAGATCAAGATAAAGAACTCGAAAAAGT 420
QY 2253 aaaaactgttcagtaactatggtcgttgattcgcgcgcaaaaaatacacacagaagcta 2312
Db 421 AAAAAGTTTCAAGTAATGCTGTGAATTCGCGCCGAAATAATACATACAGGAAGCTA 480
QY 2313 acaatgggggctacagtttgcgaattgcagctgaataaacaattagaagataaagaata 2372
Db 481 ACAATGGAGGCTTACAGTTTGCAGATTCAGCTGAATTAACCAATTAGAAGATAAAGAATA 540
QY 2373 ttgtattaaaaggagatataaattacagagaattttctttattcttt 2421
Db 541 TTTGATTAAGAGAGTGTATAAATTACAGAGAAATTTCTTTATTCTTT 589

RESULT 11
AW032987
LOCUS
563 bp mRNA linear EST 18-MAY-2001

DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

EST276546 tomato callus, TAMU Lycopersicon esculentum cDNA clone
cLEC17L14, mRNA sequence.
AW032987
AW032987.1 GI:5891743
EST
tomato.
Lycopersicon esculentum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; euasterids I; Solanales; Solanaceae; Solanum;
Lycopersicon
1 (bases 1 to 563)
Alcala,J., Vrebalov,J., White,R., Matern,A.L., Vision,T., Holt,I.E.,
Liang,F., Opton,J., Craven,M.B., Bowman,C.L., Ahn,S., Ronning
C.M., Fraser,C.M., Martin,G.B., Tanksley,S.D. and Giovannoni,J.
Generation of ESTs from tomato callus tissue
Unpublished (1999)
Contact: CUGI
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Email: <http://www.genome.clemson.edu/orders/index.html>
5 prime sequence.

FEATURES
source

Location/Qualifiers
1..563
/organism="Lycopersicon esculentum"
/cultivar="TA496"
/db_xref="taxon:4081"
/clone="cLEC17L14"
/clone_lib="tomato callus, TAMU"
/tissue_type="callus"
/dev_stage="25-40 days old"
/lab_host="XLI-Blue MRF"
/note="Vector: pBlueScript SK(-); Site.1: EcoRI; Site.2:
XhoI; supplier: Giovannoni laboratory; cLEC - Cotyledons
of seedlings 7-10 days post-germination with no selection, cut
at both ends and placed on MS medium with no selection.
Mixed callus was harvested at 25 and 40 days and included
undifferentiated masses. Tomato Callus EST Library"
159 a 106 c 135 g 163 t

Query Match 21.8%; Score 556.6; DB 9; Length 563;
Best Local Similarity 99.3%; Pred. No. 2.6e-68;
Matches 559; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 616 gtagaatgggaaattgggtgctctggttaaaagcttatatacaaaatggcgagctcaaatgct 675
Db 1 GTAGAATGGGAAATTTGGTCTCTCTCGGGAAGCTTATACAAAATGGCAGCTCAAAATGGCT 60
QY 676 gtagggttgaaaaactggtgccatggatcatgtgaagcaagagatgctctgactct 735
Db 61 GTAGGTTTGAANAACCTGGTGTCCCATGGATCATGTGTAGCAAGAGATGCTCTGATCTCT 120
QY 736 gtgattgatacttgaatggcttctactgcgaagggttcctcctaataaagccttacaaa 795
Db 121 GTGATTGATACTGTGAATGGCTTCTACTGCGAAGGTTCCGCTCTTAATAAGCTTTACAAA 180
QY 796 cctaaaaatgtggacagaagatgagctggctggtatagaaaattcgggtgggtcccaattct 855
Db 181 CCTAAAATGTGGACAGAGATATGGACTGGCTGGTATACGAAATTCGGTGGTCCAAATTCCT 240
QY 856 caaagaccagccgaagacattgcatttcagttgcagggtttgttcagaaacaatggttca 915
Db 241 CAAAGACCAGCGAAGACATTCGATTTTCAGTTGCAGGTTGTTCATAACAAATGGTTCA 300
QY 916 ttcttcaattactacatgtatcatatggaggacaacaattttggccggacatcatcagggtt 975
Db 301 TTCTTCAATTACTACATGTATCATGGAGAACAAATTTTGGCGGACATCATCAGGGCTT 360
QY 976 ttcatgcaactagctagcatttatgatctcctctcgatgaatatgggttgcgtgaatgaa 1035
|||||

```

Db 361 TTCATTGCAACAGTACGATGATGATGCTCTCGATGAATATATGGGTGCTGAATGAA 420
QY 1036 ccaaaagtggcacttgagagacttacataaagctatcaagctatctgaaccgcttta 1095
Db 421 CCAAAGTATGGGCACTTGAGAGACTTACATAAAGCTATCAAGCTATCTGAACGGGTTTA 480
QY 1096 gtttcacatctgctgcggtgactagtcttggaaagtaatacaagagctcatgtttataga 1155
Db 481 GTTTCATCATATGCTCGGCTGACTAGTCTTGAAGTAAACAAGAGGCTCATGTTTATAGA 540
QY 1156 tcaaaatctggagcttgctgctgc 1178
Db 541 TCAAAATCTGGAGCTTGTGCTGC 563

RESULT 12
BE431542
LOCUS BE431542 551 bp mRNA linear EST 18-MAY-2001
DEFINITION EST336357 tomato breaker fruit, TIGR Lycopersicon esculentum cDNA
clone cLEG6017, mRNA sequence.
ACCESSION BE431542
VERSION BE431542.1 GI:9429385
KEYWORDS EST.
SOURCE tomato.
ORGANISM Lycopersicon esculentum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; euasterids I; Solanales; Solanaceae; Solanum;
Lycopersicon.
REFERENCE 1 (bases 1 to 551)
AUTHORS Alcalá,J., Vrebalov,J., White,R., van der Hoeven,R.S., Holt,I.E.,
Liang,F., Hansen,T.S., Craven,M.B., Bowman,C.L., Ronning,C.M.,
Nierman,W., Fraser,C.M., Martin,G.B., Giovannoni,J.J. and Tanksley
,S.D.
TITLE Generation of ESTs from tomato fruit tissue, breaker stage
JOURNAL Unpublished (2000)
COMMENT Contact: CUGI
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Email: http://www.genome.clemson.edu/orders/index.html
5 prime sequence.

FEATURES
Location/Qualifiers
1..551
/organism="Lycopersicon esculentum"
/cultivar="TA496"
/db_xref="taxon:4081"
/clone_lib="tomato breaker fruit, TIGR"
/tissue_type="pericarp"
/dev_stage="breaker"
/lab_host="SOLR"
/Note="Vector: pBluescriptSKmCudapt; Site_1: EcoRI;
Site_2: XhoI; Fruit were harvested at the breaker stage
(first sign of lycopene accumulation on the blossom end of
the fruit). Fruit were cut in half and the seeds and
locules were discarded prior to freezing the pericarp."
BASE COUNT 153 a 120 c 116 g 162 t
ORIGIN

Query Match 21.5%; Score 549.4; DB 10; Length 551;
Best Local Similarity 99.8%; Pred. No. 2.6e-67;
Matches 550; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 819 gactggctgtatcacgaattcgggtggtccaatctctcaagaccgacgaagacattgc 878
Db 1 GACTGGCTGTATCACGAATTCCGGTGTCCAAATTCCTCAAGACCAGCGGAACACATTGC 60
QY 879 atttcagttgcaggtttgttcagaacaatggttcattcttcattcaattacatgatca 938
Db 61 ATTTTCAGTTGCCAGGTTGTTCAGAACCAATGGTTCATTCTTCAATTACTACATGATCA 120

```

```

QY 939 tggaggacaacattttggccgacatcatcaggcttttcattgcaactagctacagatta 998
Db 121 TGGAGGAACAAATTTTGGCCGGACATCATCAGGGCTTTTCATTGCAACTAGCTACGATTA 180
QY 999 tcatgctctctcgatgaatatgggttgctgaatgaacaaagatctggcacttgagaga 1058
Db 181 TGATGCTCTCTCGATGAATATGGGTGCTGAATGAACCAAAAGTATGGCACCTTGAGAGA 240
QY 1059 cttacataagctatacaagctatctgaaccgcttttagtttcacatcatgctgcgctgac 1118
Db 241 CTTACATAAAGCTATCAAGCTATCTGAACCGGCTTTAGTTTTCATATATGCTGCGGTGAC 300
QY 1119 tagtcttggaaagtaatacaagagctcatgtttatagaatcaaaatctggagcttgctgc 1178
Db 301 TAGTCTTGGAAAGTAATCAAGAGGCTCATGTTATAGATCAAAATCTGGAGCTTGTGCTGC 360
QY 1179 tttttatccaactatgactctagatattcgaataaagtaacacttcagaataggccata 1238
Db 361 TTTTATCCAACTATGACTCTAGATATTACAGTAAAGTACACCTTTTCAGATAGGCCATA 420
QY 1239 caatctgctccatgctccatcagcattctccgactcaaaactgcgctttacacac 1298
Db 421 CAATCTGCCCTCATGGTCCATCAGCATTTCTCCGACTCAAAACTGCCGTTTACAAAC 480
QY 1299 tgcacaggtttaactctcaagctcggagcagataaagatgacgctgcagggtggtattctc 1358
Db 481 TGCACAGGTTAACTCTCAAAAGCTCGAGATGAAGATGACGCTGCAGGTGGTGGATTGTC 540
QY 1359 ttggcagtcacat 1369
Db 541 TTGGCAGTCAT 551

RESULT 13
AW033711
LOCUS AW033711 549 bp mRNA linear EST 18-MAY-2001
DEFINITION EST277282 tomato callus, TAMU Lycopersicon esculentum cDNA clone
cLEC29B5 similar to beta-galactosidase, mRNA sequence.
ACCESSION AW033711
VERSION AW033711.1 GI:5892467
KEYWORDS EST.
SOURCE tomato.
ORGANISM Lycopersicon esculentum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; euasterids I; Solanales; Solanaceae; Solanum;
Lycopersicon.
REFERENCE 1 (bases 1 to 549)
AUTHORS Alcalá,J., Vrebalov,J., White,R., Matern,A.L., Vision,T., Holt,I.E.,
Liang,F., Upton,J., Craven,M.B., Bowman,C.L., Ahn,S., Ronning
,C.M., Fraser,C.M., Martin,G.B., Tanksley,S.D. and Giovannoni,J.
TITLE Generation of ESTs from tomato callus tissue
JOURNAL Unpublished (1999)
COMMENT Contact: CUGI
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Email: http://www.genome.clemson.edu/orders/index.html
5 prime sequence.
Location/Qualifiers
1..549
/organism="Lycopersicon esculentum"
/cultivar="TA496"
/db_xref="taxon:4081"
/clone_lib="tomato callus, TAMU"
/tissue_type="callus"
/dev_stage="25-40 days old"
/lab_host="XLI-Blue MRF"
/Note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
XhoI; supplier: Giovannoni laboratory; cLEC - Cotyledons
of seedlings 7-10 days post-germination were excised, cut
at both ends and placed on M5 medium with no selection.
FEATURES
Location/Qualifiers
1..549
/organism="Lycopersicon esculentum"
/cultivar="TA496"
/db_xref="taxon:4081"
/clone_lib="tomato callus, TAMU"
/tissue_type="callus"
/dev_stage="25-40 days old"
/lab_host="XLI-Blue MRF"
/Note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
XhoI; supplier: Giovannoni laboratory; cLEC - Cotyledons
of seedlings 7-10 days post-germination were excised, cut
at both ends and placed on M5 medium with no selection.

```


SOURCE
ORGANISM Lycopersicon esculentum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; euasterids I; Solanales; Solanaceae; Solanum;
Lycopersicon.
1 (bases 1 to 611)
Alcala, J., Vrebalov, J., White, R., Matern, A. L., Vision, T., Holt, I. E.,
Liang, F., Upton, J., Craven, M. B., Bowman, C. L., Ahn, S., Ronning,
C. M., Fraser, C. M., Martin, G. B., Tanksley, S. D. and Giovannoni, J.
Generation of ESTs from tomato callus tissue
Unpublished (1999)
Contact: CUGI
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Email: <http://www.genome.clemson.edu/orders/index.html>
5 prime sequence.

FEATURES
source
1. 611
/organism="Lycopersicon esculentum"
/cultivar="TA496"
/db_xref="taxon:4081"
/clone="cJEC6M3"
/clone_lib="tomato callus, TAMU"
/tissue_type="callus"
/dev_stage="25-40 days old"
/lab_host="X11-Blue MRF"
/note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
XhoI; supplier: Giovannoni laboratory; cLEC - Cotyledons
of seedlings 7-10 days post-germination were excised, cut
at both ends and placed on MS medium with no selection.
Mixed callus was harvested at 25 and 40 days and included
undifferentiated masses. Tomato Callus EST Library"
186 a 104 c 144 g 177 t

Query Match 20.8%; Score 532.4; DB 9; Length 611;
Best Local Similarity 90.0%; Pred. No. 5.6e-65;
Matches 610; Conservative 0; Mismatches 1; Indels 67; Gaps 1;
Qy 302 atgagccttctctggaataataatttgaggagagatgatctgttagattcatca 361
Db 1 ATGAGCCCTTCCTCGGAAATATATTTTGAAGGAAGATATGATCTGTGTAGATCATCA 60
Qy 362 aaatggtcaagaagcagcagactttatgtcaatttcacgtattggcccttacgtctgtg 421
Db 61 AAATGGTCAAAAGACAGCAGACTTTATGTCAATTTACGTATTGGCCCTTACGCTGTGCTG 120
Qy 422 aatggaacttgggggattccctgtttggctaaaataatgtgcctggtatggaatttagaa 481
Db 121 AATGGAACCTTG----- 132
Qy 482 caacaatcagccttttaaggtggctatgcaaggattttgttcagaaaaatagttcaacatga 541
Db 133 -----GGTGGCTATGCAAGGATTTGTTTCAGAAATAGTCAACATGA 173
Qy 542 tgaagtcagaaaaattgtttgaatcgaaggaggaccataattatgccccagatcacaaa 601
Db 174 TGAAGTCAGAAAATTTGTTTGAATTCGAAGGAGGACCAATAATATTATGGCCCATAGATAAA 233
Qy 602 atgagatggaccagtagaattggaaattgtgtccctggttaagcttatcaaaaatggg 661
Db 234 ATGAGTATGGACCAAGTAGAATGGAAATTTGGTCTCTGGTAAAGCTTATACAAAATGGG 293
Qy 662 cagctcaaatggctgtagttgaaactgggtgtcccatggatcatgtgtaagcaagagg 721
Db 294 CAGCTCAATGGCTGTAGGTTTGAAACTGGTGTCCCATGGATCATGTGTAAAGCAAGAGG 353
Qy 722 atgctcctgatcctgtgattgatacttgtaattggtttctactgcgaaggggttccgtccta 781
Db 354 ATGCTCCTGATCCTGTGATTGATCTTGTAAATGGCTTCTACTGCGAAGGGTTCCGTCCTTA 413

Qy 782 ataaqccttaacaaacctaaatgtggacagaagtatggactgctggtggtatatacaaaattcg 841
Db 414 ATAAACCTTTACAAACCTTAAATGTGGACAGAGATATGGACTGGCTGGTATACGAATTCG 473
Qy 842 gtggtccaatctcctcaaaagaccagccgaagacattgcatttccagttgcccaggtttgttc 901
Db 474 GTGGTCCAATTCCTCAAAAGACCAGCCGAAGACATTGCATTTTCAGTTGCCAGGTTTGTTC 533
Qy 902 agaacaatgggttcattcttcaattactacatgtatcatggaggaacaaattttggccgga 961
Db 534 AGAACAAATGGTTTCATCTTCAATTACTACATGTATCATGGAGGAACAAATTTTGGCCGGA 593
Qy 962 catcatcagggccttttca 979
Db 594 CATCATCAGGGCTTTTCA 611
Search completed: June 25, 2002, 04:37:44
Job time: 7739 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 25, 2002, 02:29:50 ; Search time 92.36 Seconds
(without alignments)
6792.421 Million cell updates/sec

Title: US-09-701-868-4
Perfect score: 2554
Sequence: 1 aaaaaagtttcaattttt.....aaaaaaaaaaaaaaaaaaaaa 2554

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

Assigned Patents_NA:
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq.*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq.*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq.*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq.*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq.*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1205.8	47.2	2628	2	US-08-696-944-1
2	1171.2	45.9	2944	2	US-08-696-944-18
3	1168.2	45.7	2945	2	US-08-687-372-2
4	67	2.6	2322	3	US-09-436-605-1
C 5	54.8	2.1	7218	1	US-08-232-463-14
C 6	54.4	2.1	8920	2	US-08-446-855A-1
C 7	54.4	2.1	8920	4	US-09-150-741-1
8	48.8	1.9	19124	2	US-08-487-826B-13
9	47.4	1.9	3057	1	US-08-551-459-3
10	46.8	1.8	1493	6	5340934-5
11	46.2	1.8	3871	2	US-08-599-455B-3
12	46.2	1.8	3871	4	US-09-069-781B-3
C 13	45.6	1.8	470	4	US-09-020-956-102
C 14	45.6	1.8	470	4	US-09-030-607-102
C 15	45.6	1.8	470	4	US-09-439-313-102
16	45.2	1.8	3515	1	US-08-596-985-1
17	45	1.8	1641	1	US-08-300-903A-8
18	45	1.8	2445	6	5215909-9
19	44.4	1.7	1160	3	US-08-995-159-1
20	43.8	1.7	2110	4	US-09-419-459-1
21	43.8	1.7	3466	1	US-08-551-459-5
22	43.4	1.7	1240	1	US-08-248-466B-6
23	43.4	1.7	1510	1	US-08-248-466B-3
C 24	43.4	1.7	5852	1	US-07-867-106-2
25	43.4	1.7	8920	2	US-08-446-855A-1
26	43.4	1.7	8920	4	US-09-150-741-1
C 27	43.2	1.7	509	4	US-09-030-607-202

C 28	43.2	1.7	509	4	US-09-439-313-202
C 29	43	1.7	1431	4	US-09-316-083-2
C 30	43	1.7	3527	2	US-08-909-965C-7
31	42.4	1.7	1172	1	US-07-945-288-9
32	42.4	1.7	1172	1	US-08-462-831-9
33	42.4	1.7	1172	1	US-08-461-809-9
34	42.4	1.7	1172	1	US-08-461-441-9
35	42.4	1.7	1172	5	PCT-US93-08518-9
36	42.4	1.7	2550	6	5258287-23
37	42.4	1.7	6243	2	US-09-056-075-1
C 38	42.4	1.7	19124	2	US-08-487-826B-13
C 39	42.2	1.7	857	1	US-08-308-883-1
C 40	42.2	1.7	857	1	US-08-730-163-1
C 41	42.2	1.7	857	4	US-08-256-799-1
C 42	42.2	1.7	857	4	US-08-462-437-1
C 43	42.2	1.7	2058	2	US-08-749-391-1
44	42.2	1.7	2058	3	US-09-390-200-1
45	42.2	1.7	2322	1	US-08-618-164-1

ALIGNMENTS

RESULT 1
US-08-696-944-1
; Sequence 1, Application US/08696944
; Patent No. 5981831
; GENERAL INFORMATION:
; APPLICANT: Sumant CHENGAPPA
; APPLICANT: Susan A. HELLYER
; APPLICANT: John S. REID
; APPLICANT: Jacqueline DE SILVA
; TITLE OF INVENTION: No. 5981831el Exo-(1-4)-Beta-D Galactanase
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESS: Pillsbury Madison & Sutro, L.L.P.
; STREET: 1100 New York Avenue, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: MS Word
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/696,944
; FILING DATE: 23-AUG-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/GB95/00372
; FILING DATE: 23-FEB-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9403423.8
; FILING DATE: 23-FEB-1994
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2628 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 130..2319
; US-08-696-944-1

Query Match 47.2%; Score 1205.8; DB 2; Length 2628;
Best Local Similarity 73.5%; Pred. No. 2.9e-268;
Matches 1569; Conservative 0; Mismatches 557; Indels 9; Gaps 2;

Qy	98	tttgtttattggatttttttttcttcagtgtaaaagctagtgcttctcttatgatgacagagacta	157
Db	194		
		tggttatttggttttttttggggtttggttatgtcacagccctctgtttactttatgatcataaagcca	253
Qy	158	taatcaataatgggaaagaaaaactctattcttcttggttcaattcatctatccaagaagca	217
Db	254		
		ttatgatattaaaggcagagaagaatttttgatctcttggtttccatttctacttccaagaagca	313
Qy	218	ctccacagatgtagcctgatctcttatacaaaaagctaaagatggagccttagatgtattg	277
Db	314		
		caacctcagatgtgccagacccttatttcanaagcccaagatgsgagggcttgatgttatag	373
Qy	278	aaactatgttttctggaaatggacatbgagcctctctcttggaataataatttgaagaa	337
Db	374		
		agactttatgtgttctggaatggacatgaaccttctcttgaaaaattattattttgaggata	433
Qy	338	gatatgatctgttagatctcaataatggtcaaaagcagcagacttatgtatcaatttac	397
Db	434		
		ggtttgacctgtgtggggttcataaaagtggttcagacagctggctatttttcttatctctca	493
Qy	398	gtattggcccttaagctctgtctgtaatggaactttgggggattccctgtttggctaaaaat	457
Db	494		
		ggatttggtcctttcatatgtgtctgaatggaactttggaggatttctgtttggctcaaat	553
Qy	458	atgtcgtcgtgataaggaaatttagaacaacaatacagccttttaagtggtgctatgcaagat	517
Db	554		
		atggtttccttggtatttgctttcagaaacagacaatgagcctttcaaggagccaatgcaaaaat	613
Qy	518	ttgttcgaataatagttcaacatgatgaagtcaaaaaattgtttgaatctcaagagagac	577
Db	614		
		tcactgagaagattgttaatatgaagaaacagagaagttgtttcattcccaggagagtc	673
Qy	578	caataattatggccagatatacaaaatgagtatggaccagtagaaatgggaataattggtgctc	637
Db	674		
		caataattctgtctcagatagagaatgactatgaccagctgggaatggaaaattggtgctc	733
Qy	638	ctggtaagctttatacaaaatggcagctcaaatggctgtagggtttgaaaactggtgtcc	697
Db	734		
		ctgaaaagctttataccaaatgggctgcttcaaatggctgtaggctcagatactgggtgcc	793
Qy	698	catgatcatgtataagcaagagatgctccttgatctcttgattgatacttgtaatggct	757
Db	794		
		catggggttatgtccagcagaagaatgcactttgatctcttattttatgtatacctgcaatggat	853
Qy	758	tctactgcgaagggttcctctctaataagccttatacaacctaaatgtggacagaagatat	817
Db	854		
		ttttactgtgaaaaacttctactccaaaacaagaactacaaaacccaaaattgtggacagaaaaat	913
Qy	818	ggactggctggtatagaaattcgggtgttcgaattctctcaagaccagccgaagaacattg	877
Db	914		
		ggactggctgggtacactgctttttgggtgggtgcAACCCCTTATAGACCACGACAAGATATAG	973
Qy	878	cattttcagttgcccaggtttgttcagaacaaatggttcattcttcaatcacatcatgctc	937
Db	974		
		caittttcagttggccagatttcatttcagatcgcggctctactcttttaactacttatgtatc	1033
Qy	938	atggaggaacaaattttggccggacatcatcagggtcttttcattgcaactagctacgatt	997
Db	1034		
		atggaggaacttaacttttggccggacatcgaatggcctctcttggtgccacaagtattgact	1093
Qy	998	atgatgctcctctcgatgaatatgggttgctgaatgaacccaaagtaatgggcacttgagag	1057
Db	1094		
		atgatgctcccatattgatgaatatggagacttttcaatgaacaaaaatggggggcatctgagag	1153
Qy	1058	acttacataaagctataagctatctgaaccggcctttagtttcaatcatatgctcgggtga	1117
Db	1154		
		aatttacatagagcaataaaaacaatgcgagctgcgcttttagtgcgtggatccccacagtgt	1213
Qy	1118	ctagctcttggaaagtaatcaagaggtctatgtttatagatcaaaatctggagcttgtgctg	1177
Db	1214	catggccttgaaaaaacctttgaggtgacatttctgataacacagagagctctctgctgctg	1270

5

2

Db	531	ACTAAGATTCCTTGATATGATGAAGCAGAAAAAGCTCTATGAACACTCAGGGGTGGTCCAATT	590
Qy	583	attatggcccgatatacaaaatgagtatggacagtagaatgggaaattgggtgctcctgg	642
Db	591	ATTCTATCTCAGATAGAAAATGAATGACCTATGGAGTGGGAACACTAGGTGAACCTGGT	650
Qy	643	aaagcttatatacaaaatgggcagctcaaatggctgtaggtttgaaacctgggtgctcccatgg	702
Db	651	AAAGTTTACTCAGAATGGCGACCCAAAATGGCTGGGAATCTTGGCACTGGTGTCCCATGG	710
Qy	703	atcatgtgaagcaagagatgctcctgaatcctgtagtctgattgatacttgaatggctcttac	762
Db	711	ATCATGTGCAGCAAGATGATGCTCGTATCTTATTAATACTTGAACATGGTTTCTTAC	770
Qy	763	tgcgaagggttcgcgtcctaataagccttacaacctaaaatgtggacagaagtatggact	822
Db	771	TGTGACTACTTTCACCAAAATGAAGCTAATAAACCCCAAGATGTGSACTGAAGCCTGGACA	830
Qy	823	ggctgggtatacgaaaatcggtgggtccaaatctctcaaaagccagccgaagacattgcattt	882
Db	831	GCCTGGTTTACCGAATTTGGAGGTCAGTTCCTTACCCTCTCGCAGAGGATATGGCATTT	890
Qy	883	tcagttgcaggtttgttcagaaacaatggctccattcttcaattacttacatcatgatacgg	942
Db	891	CTGTGCGAAGATTTTATACAAACGGGAGGCTCCTTCAATTAATTAATATGATCATGG	950
Qy	943	ggacaataattggccggacatcatcagggcttttcattgcaactagctagcatgattatgat	1002
Db	951	GGACAAACTTTGGAAGGACTTCTGGTGGCCCATTTATTTGCTACTAGTATGATATATGAT	1010
Qy	1003	gtcctctcgatgaatatgggtgtcgtgaatgaaccaaagtatggcacttgagagactta	1062
Db	1011	GCACCCTAGATGAATTTGGGTCAATTCGGCAGCCCTAAATGGGGTCATCTGAAGAATCTA	1070
Qy	1063	cataagctatcaagctatctgaacgcgcttagtttcatcatatgctgctgggtgaactagt	1122
Db	1071	CATAGAGCAATAAAGCTCTGTGAGCAGCCTTAGTATCTGTAGATCCAACCTGTGACATCC	1130
Qy	1123	cttgggaagtaatacaagagctcatgtttatagatacaaaatctgagcttctgctgctttt	1182
Db	1131	TTAGGAAACTATCAGAGGCAGCTGTTTTCAGTCAGAGTCCTGGGGCCTGGGCTGCCCTTC	1190
Qy	1183	tatccaaactatgactctagatatctcagtaaaagtcacctttcagaataatgagccatacaat	1242
Db	1191	CTAGCAAAATTACACCAGCACTCTTTTGGTAAAGTGGCATTTTGGGAACATGCAATTATAAC	1250
Qy	1243	ctgctctcaatggctcaatcaagaattctccgagctgcaaaaactgcgctttacaacactgca	1302
Db	1251	TTGCCACCTGGTCTATCAGCAATCTCCGCACTGCAAGAACACTGTCTATAATCTAGTCA	1310
Qy	1303	caggttcaactctcaagctcagagcaataagatgacgcctgcaggtggtgatctctctgg	1362
Db	1311	AGGGTTGGTCTCAAGTGCTCAGATGAGATGACTCCAGTCAGTAGAGATTCATGAGTCT	1370
Qy	1363	cagtcatacaatgaagaaaacgctactctgatgacgcgatacaacttacagctaaacgga	1422
Db	1371	GAGTCATTAATGAAGACGACGATCGCATGAAGA--CGACACTTTCACAGTTGTGTGGG	1427
Qy	1423	ctatgggaacagaaaaacgtcacagagatttcacagactatctgtgttgatagacaat	1482
Db	1428	TTATGGAGCAGATTAATATACAGAGATGTATCTGATTACTTGTGTGTATATGACTGAC	1487
Qy	1483	gtaaatatgacatcctaataagaatttctaagaacggaaaaggatccttatctcaactgtt	1542
Db	1488	ATTGAGATTGATCCAACAGAGGATTTTGAATAGTGAAATTTGGCCTTGCTTACTGTCT	1547
Qy	1543	atgctccgtggctcatgtcttgcattgttcgttcaatgaaaactatcaggaactgttat	1602
Db	1548	TTCTCTGCTGCCATGTCATGTATCTGTGTAATGGTCAATTTAGCAGGAACGTGTGTAC	1607
Qy	1603	ggtaacattggataatccaaaacttacatcacagtggcaacgtgaagtttaagagctggatt	1662
Db	1608	GGAAGTTTGAAGAACCCAAACTAACTTTTCAGAACGGTATTAATCTTCAGAGCTGGGTG	1667

RESULT 4

US-09-436-605-1

; Sequence 1, Application US/09436605

; Patent No. 6140115

; GENERAL INFORMATION:

; APPLICANT: Kolodny, Edwin, H.

APPLICANT: Wang, Zhao-Hui

APPLICANT: Raghavan, Srinivasa,

: APPLICANT: Seng. Baijin

TITLE OF INVENTION: The Canine (-Galactosidase Gene and Gm1-Gangliosidosis

: FILE REFERENCE: D6273

FILE REFERENCE: D08273
CURRENT APPLICATION NUMBER: US/09/436 605

; CURRENT APPLICATION NUMBER: US/
 : CURRENT FILING DATE: 1999-11-09

;; CURRENT FILING DATE: 1999
; NUMBER OF SEC ID NOS: 25

; NUMBER OF
; CEO ID NO 1

```

; SEQ ID NO 1
: LENGTH: 333

```

; LENGTH: 2322

; TYPE: DNA

; ORGANISM: Can

```

; FEATURE:

```

; NAME/KEY: CDS

```

; LOCATION: 22..2028

```

; OTHER INFORMATION: Contains coding sequence for Portuguese Water dog

OTHER INFORMATION: (-Galactosidase. A g200 to a200 substitution mut

OTHER INFORMATION: is present in the a

OTHER INFORMATION


```

; FILING DATE: 06-Jul-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Mitchard, Leonard C
; REGISTRATION NUMBER: 29,009
; REFERENCE/DOCKET NUMBER: 47-80
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-816-4000
; TELEFAX: 703-816-4100
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8920 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: genomic
; US-08-446-855A-1

Query Match 2.1%; Score 54.4; DB 2; Length 8920;
Best Local Similarity 49.4%; Pred. No. 0.0018;
Matches 168; Conservative 0; Mismatches 171; Indels 1; Gaps 1;

QY 2212 attctctagtcagagatcaagataaagaactcgaaaagtaaaactgttcagtaacta 2271
Db 875 ATATATATACATTTTAATAACAACACTTGTGATGTTAAAGAAATAAAAGCTGTTTAAAGACCTA 816
QY 2272 tggctgctgaattcgccgcgcaaaaatacatcacgaaagctaacatggagcgctacagttt 2331
Db 815 TGATTCAGAGAAATATCCCAATATATATATATATATATATATATATATATATATATATATAT 756
QY 2332 gcaaatgacgagctgaataaacaattagaaagataaagaataattgttaaaggaggtata 2391
Db 755 TATTTTTCCTCCATTTTCTTTTTTTTTTTTATACATTTATATATTTATGTTTAAATATTTTA 696
QY 2392 taaattttacagaga-atttcttctattctttgttaaaacttttggtttataaagttttataca 2450
Db 695 TAAATTTTACATATACAGAGTTCATTTTTCATATGTAAGTAAATTTTTTTTTTTCTTTTTTTT 636
QY 2451 gaatttctgttattggattgatgattgagaagattgtcacagctcccaataactatt 2510
Db 635 TTTTTTTTTTTTTTTTTTTTTTTTAAATTTAGTAGAATTTACTATTTTATATAAATAAGAAAAAAT 576
QY 2511 agaatacaaataaatttctatgttaaaaaaataaaaaa 2550
Db 575 AAATAAATGAATAAAATTAATAAATAAATAATATATATAAATAAAA 536

RESULT 7
US-09-150-741-1/c
; Sequence 1, Application US/09150741
; Patent No. 6183996
; GENERAL INFORMATION:
; APPLICANT: Stewart et al.
; TITLE OF INVENTION: Nucleotide Sequence Encoding Carbamoyl Phosphate
; Patent No. 6183996
; TITLE OF INVENTION: Synthetase II
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/09/150,741
; CURRENT FILING DATE: 1998-09-10
; EARLIER APPLICATION NUMBER: PL6380
; EARLIER FILING DATE: 1992-12-16
; EARLIER APPLICATION NUMBER: AU93/00617
; EARLIER FILING DATE: 1993-12-02
; EARLIER APPLICATION NUMBER: 08/446,855
; EARLIER FILING DATE: 1995-07-06
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 8920
; TYPE: DNA
; ORGANISM: Plasmodium falciparum
US-09-150-741-1

```


THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 25, 2002, 03:33:55 ; Search time 489.19 Seconds
(without alignments)
8963.796 Million cell updates/sec

Title: US-09-701-868-4
Perfect score: 2554
Sequence: 1 aaaaaagtttcaattttt.....aaaaaaaaaaaaaaaaaaaaa 2554

Scoring table: IDENTITY_NUC
Gap 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues
Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : ANGenes032802
1: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1980.DAT.*
2: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1981.DAT.*
3: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1982.DAT.*
4: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1983.DAT.*
5: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1984.DAT.*
6: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1985.DAT.*
7: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1986.DAT.*
8: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1987.DAT.*
9: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1988.DAT.*
10: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1989.DAT.*
11: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1990.DAT.*
12: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1991.DAT.*
13: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1992.DAT.*
14: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1993.DAT.*
15: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1994.DAT.*
16: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1995.DAT.*
17: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1996.DAT.*
18: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1997.DAT.*
19: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1998.DAT.*
20: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2000.DAT.*
21: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT.*
22: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001B.DAT.*
23: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2002.DAT.*
24: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	ID	Description
1	2546	99.7	2554	21	AAZ29341 Tomato beta galact
2	1205.8	47.2	2628	16	AAZ01014 Lupin exo-(1-4)bet
3	1172.8	45.9	3069	21	AAZ29340 Tomato beta galact
4	1171.2	45.9	2944	16	AAZ01015 Tomato exo-(1-4)be
5	1168.2	45.7	2945	16	AAQ89602 Tomato beta galact
6	1168.2	45.7	3224	21	AAZ29338 Tomato beta galact
7	1108	43.4	2469	21	AAC49769 Arabidopsis thalia
8	952.4	37.3	2571	21	AAC49810 Arabidopsis thalia
9	877	34.3	2553	21	AAZ46270 Arabidopsis thalia

10	877	34.3	2562	21	AAZ42925 Arabidopsis thalia
11	708.8	27.8	2778	17	AAZ15173 Nasturtium xyloglu
12	649.4	25.4	2972	21	AAZ29344 Tomato beta galact
13	507.8	19.9	2984	21	AAZ29339 Tomato beta galact
14	388.4	15.2	749	21	AAZ29343 Tomato beta galact
15	372.6	14.6	1969	21	AAZ40074 Arabidopsis thalia
16	331.4	13.0	755	21	AAZ29342 Tomato beta galact
17	242.8	9.5	890	21	AAZ41630 Arabidopsis thalia
18	231.2	9.1	1240	21	AAZ38606 Arabidopsis thalia
19	208.6	8.2	458	21	AAZ36838 Arabidopsis thalia
20	152	6.0	457	21	AAZ44756 Arabidopsis thalia
c	102.4	4.0	8546	20	AAZ13083 Enterococcus faeca
21	96	3.8	8991	20	AAZ13195 Enterococcus faeca
22	94	3.7	14736	19	AAV52304 Streptococcus pneu
23	91.4	3.6	542	21	AAC95011 Cat flea hindgut a
24	86.8	3.4	832	17	AAZ37082 AR782 fragment enc
25	76.6	3.0	6279	20	AAZ13280 Enterococcus faeca
26	75.6	3.0	513445	22	AAI61373 Soybean 318013 reg
27	73	2.9	3030	22	AAH42268 Nucleotide sequenc
28	72.8	2.9	3039	22	AAK94894 Human full-length
29	72.2	2.8	1908	21	AAAG2060 Hydrophobic domain
30	72.2	2.8	3024	22	AAK94298 Human full-length
31	72.2	2.8	3106	20	AAZ52245 Protein PRO236 CDN
32	72.2	2.8	3106	22	AAZ52245 Human DNA encoding
33	72.2	2.8	3106	22	AAZ52245 Human PRO236 cDNA
34	72.2	2.8	3234	21	AAAG2070 Hydrophobic domain
35	72.2	2.8	819	22	AAK91999 Human cDNA 5'-end
36	70.6	2.8	819	22	AAK93676 Human cDNA clone r
37	70.6	2.8	2906	22	AAK94350 Human full-length
38	68.4	2.7	434	21	AAZ52906 Arabidopsis thalia
39	67	2.6	2322	22	AAZ52906 Portuguese Water d
40	65.2	2.6	440	21	AAZ52906 Arabidopsis thalia
41	63.8	2.5	1794	20	AAZ88791 Lactobacillus case
42	63.8	2.5	2409	19	AAZ42728 Human beta galacto
43	62.2	2.4	2435	21	AAZ15878 Human prostate can
44	61.8	2.4	669	22	AAZ32666 Human cDNA encodin
45					

ALIGNMENTS

RESULT	1
AAZ29341	
ID	AAZ29341 standard; cDNA; 2554 BP.
XX	AC AAZ29341;
XX	AC AAZ29341;
DT	29-FEB-2000 (first entry)
XX	Tomato beta galactosidase-4 cDNA.
DE	Tomato beta galactosidase-4; TBG-4; clone pZBG2-1-4; fruit ripening;
KW	Rutgers Tomato plant; pectin; biofilm; gene mapping; ds.
XX	Lycopersicon esculentum.
OS	Lycopersicon esculentum.
XX	Key Location/Qualifiers
FT	CDS 64..2238
FT	/*tag=
FT	/product= "Tomato beta galactosidase-4"
XX	WO9964564-A1.
PN	16-DEC-1999.
PD	08-JUN-1999; 99WO-US12697.
XX	09-JUN-1998; 98US-0088805.
PR	(USDA) US DEPT OF AGRICULTURE.
XX	PI Gross KC, Smith DL;
XX	

DR WPI; 2000-097532/08.
DR P-PSDB; AAY44306.

XX New beta-galactosidases, used to prepare transgenic plants with altered fruit ripening -

PS Claim 2; Fig 2; 85pp; English.

The present sequence encodes tomato beta galactosidase-4 (TBG-4). This is a cDNA insert of clone pZBG2-1-4 from a cDNA library derived from breaker, turning and pink fruit pericarp from 'Rutgers' tomato plants. TBG cDNA has been expressed in *E. coli* strain Xlilblue MR. This is used for modifying cell wall metabolism and controlling ripening of fruit by altering activity of beta galactosidase II protein. Pectin with reduced galactosyl content is produced for use in biofilms or solutions. Fragments of the cDNA are used for gene mapping and for detecting expression of beta galactosidase in plant tissue. Transgenic plants with altered fruit ripening are produced by introducing DNA constructs comprising a TBG cDNA associated with promoter/enhancer elements.

Sequence 2554 BP; 822 A; 431 C; 559 G; 742 T; 0 other;

Query Match	99.7%	Score 2546;	DB 21;	Length 2554;
Best Local Similarity	99.8%	Pred. No. 0;		
Matches 2549;	Conservative	0;	Mismatches 5;	Indels 0;
Gaps 0;				

[illegible]


```
Oy 1801 tgggttacaaggttggtcgtgaaagcgaaatcggttaagtcttcaactccttaagtggagt 1860
    |||
Db 1801 tgggttacaaggttggtcgtgaaagcgaaatcggttaagtcttcaactccttaagtggagt 1860
Oy 1861 tcttctgttgatgggttcgaggttcacataatggtctcaaaagcagccctgacttgggtac 1920
    |||
Db 1861 tcttctgttgatgggttcgaggttcacataatggtctcaaaagcagccctgacttgggtac 1920
Oy 1921 aaggtctacattaaacgagccctggaggaaatgatccactagcttttagacatggcaagtatg 1980
    |||
Db 1921 aaggtctacattaaacgagccctggaggaaatgatccactagcttttagacatggcaagtatg 1980
Oy 1981 ggaagaggtcagatgatgataaaatgggtgaagcgtagtgcgaccttggcctggatcacata 2040
    |||
Db 1981 ggaagaggtcagatgatgataaaatgggtgaagcgtagtgcgaccttggcctggatcacata 2040
Oy 2041 gcacaaagcgactgcagcaaatgaagtattatgctgaaacgttcaacagagaagaatgcgga 2100
    |||
Db 2041 gcacaaagcgactgcagcaaatgaagtattatgctgaaacgttcaacagagaagaatgcgga 2100
Oy 2101 actaactgcgacaacctctcagagatggttaccatgttccacgatcgtgctgaaaccca 2160
    |||
Db 2101 actaactgcgacaacctctcagagatggttaccatgttccacgatcgtgctgaaaccca 2160
Oy 2161 agtggaactttagtagtattcgaagaatgggaggtgaatccacaggaatttctcta 2220
    |||
Db 2161 agtggaactttagtagtattcgaagaatgggaggtgaatccacaggaatttctcta 2220
Oy 2221 gtcaggagatcaagataaagaactgcgaagaatgaagaactgttcagtaactatggtgcttg 2280
    |||
Db 2221 gtcaggagatcaagataaagaactgcgaagaatgaagaactgttcagtaactatggtgcttg 2280
Oy 2281 aattcgcgcgaataatcatatcacgaagctaaacatgagagctacagtttgcaaatgac 2340
    |||
Db 2281 aattcgcgcgaataatcatatcacgaagctaaacatgagagctacagtttgcaaatgac 2340
Oy 2341 agctgaataaaacattagaagataaaagaataatttggataaaaggagtataaaatttac 2400
    |||
Db 2341 agctgaataaaacattagaagataaaagaataatttggataaaaggagtataaaatttac 2400
Oy 2401 agagaatttcttattcttctgtaaaacttgggtttataaaagtttatcacagaatttctg 2460
    |||
Db 2401 agagaatttcttattcttctgtaaaacttgggtttataaaagtttatcacagaatttctg 2460
Oy 2461 ttatttgattatgagattgagaagattgtacagcttccaaactactattagatcacaaa 2520
    |||
Db 2461 ttatttgattatgagattgagaagattgtacagcttccaaactactattagatcacaaa 2520
Oy 2521 taaatttcatgtaaaaaaaataaaaaaaataaaaaa 2554
    |||
Db 2521 taaatttcatgtaaaaaaaataaaaaaaataaaaaa 2554
```

RESULT 2

```
AAT01014
ID AAT01014 standard; cDNA; 2628 BP.
```

XX AC AAT01014;

XX XX

XX DT 10-FEB-1996 (first entry)

XX XX

DE Lupin exo-(1-4)-beta-D-galactanase cDNA sequence.

```
KW Exo-(1-4)-beta-D-galactanase; enzyme; transgenic plant;
KW crop improvement; ss.
```

XX Lupus angustifolius.

OS

XX Key Location/Qualifiers

FT CDS 130..2322

FT /*tag= a

XX XX

```
PN WO9523228-A1.
XX 31-AUG-1995.
XX 23-FEB-1995; 95WO-GB00372.
XX 23-FEB-1994; 94GB-0003423.
XX (UNIL ) UNILEVER NV.
PA (UNIL ) UNILEVER PLC.
XX Chengappa S, De SILVA J, Hellyer SA, Reid JSG;
PI WPI; 1995-311537/40.
XX P-PSDB; AAR82881.
DR Lupin and tomato exo-galactanase DNA and protein - useful for the
XX alteration of plant characteristic(s), e.g. texture, growth,
XX ripening
PS Claim 1; Page 25; 67pp; English.
XX CC This cDNA sequence may be used in the construction of a transgenic
XX plant in order to alter characteristics of e.g. growth, texture
XX or ripening of the plant or plant parts. This sequence may also be
XX expressed in a recombinant host for the production of the enzyme
XX which may be used for the modification, degradation or liquefaction
XX of plant materials in order to affect mechanical properties relating
XX to eating texture, particle sizes of, e.g. fruit or vegetable juices,
XX or extractability of colors, flavors or vitamins.
SQ Sequence 2628 BP; 835 A; 448 C; 581 G; 764 T; 0 other;
```

```
Query_Match 47.2%; Score 1205.8; DB 16; Length 2628;
Best Local Similarity 73.5%; Pred. No. 5.7e-234;
Matches 1569; Conservative 0; Mismatches 557; Indels 9; Gaps 2;
```

```
Oy 98 ttgtttattggtattttttcttcagtgaaagctagtttcttctatgatgacagacta 157
    |||
Db 194 tgtattgttttttgggtttgttatgtcacagctctgttacttatgatcataaagcca 253
Oy 158 taatcataaatgggaaaaaataattcttatttctgttcaattcattatccaagaagca 217
    |||
Db 254 ttatgattaaaggcgagagaagaatttgcctctgttccattcactatccaagaagca 313
Oy 218 ctcaacagatgtgacctgatcttatacaaaagctcaagatgagagcttagatgtattg 277
    |||
Db 314 cacctcagatgtggccagaccttattcaaaaggccaaagatggagggttgatgtattag 373
Oy 278 aaacttatgtttctggagtgcacatgagccttctccttgaaaaataataatttgaagaa 337
    |||
Db 374 agacttatgtctctggaaatggacatgaaccttctccttgaaaaataattatttggaggata 433
Oy 338 gatatgatctttagattcatcaaaatgggtacaaagcagcaggaactttatgtcaatttac 397
    |||
Db 434 ggttgacctgtgtgggttcataaaagtgttcagcaagctgcttatttgtctatcctca 493
Oy 398 gtattggcccttaacgtctgtgctgaatggaaacttgggggattccctgttggctaaaaat 457
    |||
Db 494 ggtattgctcttcattatgtctgaatggaaacttggaggatttctctgttggcctcaat 553
Oy 458 atgtgctcgttgatggaatttagaacaacaatcacgcttcttaagtggctgcaagat 517
    |||
Db 554 atgttctcgttgatttcttcagaacaacatgagccttccaaggaggaatgcaaaaaat 613
Oy 518 ttgttcagaaaaatagtcacatgatgaatgcagaaaaatttgtttgaattctcaaggagac 577
    |||
Db 614 tcactgagaagattgtataataatgaagcagagaagtgttccaatccccaggaggtc 673
Oy 578 caataattatggcccatgatacaaaatgatgtgaccagtagaattgggaaattggtgtc 637
    |||
Db 674 caataattctgtctcagatagagaatagtagtgaccagtggaatgggaaattggtgtc 733
```

QY 638 ctggtaaaagcttatcaaaaatgggcagctcaaatggctgtaggtttgaaacatggtgtcc 697
Db 734 ctggaagcttatcaaaaatgggcagctcaaatggctgtaggtttgaaacatggtgtcc 793
QY 698 catggtatctgtgaagaagaagatgctcctgacccctggtgattgatacttgaatgctt 757
Db 794 catgggttatgtgaagaagaagatgctcctgacccctggtgattgatacttgaatgctt 853
QY 758 tctactgcgaaggggttcgcctctaataagccttacaacacctaataatgtggacagaagt 817
Db 854 ttactgtgaaaacttcaactcaacaagaactacaacacctaataatgtggacagaagt 913
QY 818 ggaatggctgtatacgaataatgggtggtccaaattctccaaagaccagccgaagacattg 877
Db 914 ggaatggctgtatacgaataatgggtggtccaaacttatagaccagcagaagatatag 973
QY 878 catcttcagttgcaggtttgttcagaacaatgggttcattcttcattactacatgatac 937
Db 974 catcttcagttgcaggtttgttcagaacaatgggttcattcttcattactacatgatac 1033
QY 938 atggaggacaataatttggccgacatcatcagggcttttcatgtcaactagctacgatt 997
Db 1034 atggaggacaataatttggccgacatcagaaatgggtctctctgttcccaaaagtattgact 1093
QY 998 atgagtctctctcagatgaataatgggtgtgctgaatgaacccaagaatggtggcacttgag 1057
Db 1094 atgagtctctctcagatgaataatgggtgtgctgaatgaacccaagaatggtggcacttgag 1153
QY 1058 acttcataaaagctatcagctatgataacccggtttgatttcacatcattgctggtga 1117
Db 1154 aattcattagagcaataaacaatcgagtcggtttgatttcacatcattgctggtga 1213
QY 1118 ctgactcttgaaagtaatacaagggctcatgtttatagatcaaaatctggagcttgcgtg 1177
Db 1214 catggcttgaaacacttgaggtacattgttacaagaacagagctc---gctgtgctg 1270
QY 1178 cttttttatccaaactgactctagatattcagtaaaagtcaccccttcagaatagccat 1237
Db 1271 catcttctgcaattataacacccgactattcaacgaagttaaatcggaaatggacaat 1330
QY 1238 acaatctgcctccatggtccatcagcattcttccgactgcaaaactccgctttacaaca 1297
Db 1331 atgactcaccacttggctatcagattcttccgactgcaaaactgaagttttacaaca 1390
QY 1298 ctgcacaggttaactctcaaaagctcgagcataaagatgacgctcgaggtggtgattct 1357
Db 1391 ctgcaaggttaattcccgagattacatagaaatgactccagtaaacagtgatttg 1450
QY 1358 ctggcagtcatacaatgaagaaacgctactgctgatgacagcgcgatacaacttacagcta 1417
Db 1451 ctggcagtcatacaatgaagaaacgctactgctgatgacagcgcgatacaacttacagcta 1510
QY 1418 accgactatgggaacagaaaaacgtcacaaagattcaccagactatctgtgtacatga 1477
Db 1511 atgactatgggagcaggttggtgagcccgcatcttccgactgcaaaactgaagttttaca 1570
QY 1478 caaatgtaaatatagcatcctaatgaagattctcaaaagcgaaggtacacttactca 1537
Db 1571 cagatgtcaacaattggtcctaataga-----tataaagatgggaaatggcctgttctga 1624
QY 1538 ctggttatgtccgctggtcattcttgcattgttttgcataatggaaacatacaggaactg 1597
Db 1625 cagcaatgtcagcaggtcattcttgcattgttttgcataatggaaacatacaggaactg 1684
QY 1598 ttatgtgtacattgtgataatcaaacattacatacagtggaacgtgaagttaaagactg 1657
Db 1685 catatggaggtcagatcatcagatcaaacatttagtcaaaagtggaactcagagattg 1744
QY 1658 gtatacaacaagattctctgctcaggttttccgttggctcccgaaactggtgctgcaatt 1717
Db 1745 gcaatacaacaagattcttacttagtttccgttggctcccgaaactggtgctcact 1804

QY 1718 atgatacatggaatgcagaggttcttaggtccagtcacgttgagcgttccaatgaaggt 1777
Db 1805 ttgagacatggaatgactgaggtgtgtgtccagtcacactgacaggttccaatgaaggt 1864
QY 1778 caagaaacttggcgaacagaaatggtcttacaaggttggctgaaagcgaatgtaa 1837
Db 1865 catgggtattcttcgaagcaaaatggtcttacaaggttggctgaaagcgaatgtaa 1924
QY 1838 gcttccactccttaagtggtgagttctctgttgaatgggttcgaggttcaactaaaggctc 1897
Db 1925 gcttccactccttaagtggtgagttctctgttgaatgggttcgaggttcaactaaaggctc 1984
QY 1898 aaaaagacccctgacttgggtacaaggtcacaattttaacgcgcctggaggaataatgatccac 1957
Db 1985 aaaaacaccccttggcatggtataagcaaaccttttagcgacccgcaacgacgactcgt 2044
QY 1958 tagctttagacatgccaagtatgggaaggtcgaaggtcagatatgataaatggtgaagcgtag 2017
Db 2045 tggcctcgtgatttagtgatcgttgtaaggggtgaagtatggtaaatggtcacaagcattg 2104
QY 2018 gtgcctcattggcctgatacatagcacaaagcgactgcagcaaatgcagttatgctggaa 2077
Db 2105 gacgcaattggcctgaaataaagctcgtgtaattgcgcaattgtaattacgctggaa 2164
QY 2078 ogttcaacgagaagaatgctggaactaactcgcggaacacttctcagagatggtaccatg 2137
Db 2165 cttataccgatacaaaaatgcttagcaaaccttggacaacccctcccaagatggtatcgt 2224
QY 2138 ttccacgactcgtgctgaaacaaagtgaaacttggaaactttagtagtattcgaagaatgggag 2197
Db 2225 ttccctcgtcattggctgagatcgggtgtaacttctgtagtattgtagtattcgaagaatgggag 2284
QY 2198 gtaatcccaacaggaattctctctagtcaggagatca 2232
Db 2285 gtgacctaatggaattgcttgggtggaaagaaca 2319
RESULT 3
AAZ29340
ID AAZ29340 standard; cDNA; 3069 BP.
XX
AC AAZ29340;
XX 29-FEB-2000 (first entry)
XX Tomato beta galactosidase-3 cDNA.
XX
XX
KW Tomato beta galactosidase-3; TBG-3; clone pZBG2-1-4; fruit ripening;
KW Rutgers tomato plant; pectin; biofilm; gene mapping; ds.
XX
OS Lycopersicon esculentum.
XX
XX
FH Key Location/Qualifiers
FT CDS 122..2638
FT /*tag= a
FT /product= "Tomato beta galactosidase-3"
FT sig_peptide 122..187
FT /*tag= b
FT mat_peptide 188..2635
FT /*tag= c
XX
XX WO9964564-A1.
XX
XX 16-DEC-1999.
XX
XX 08-JUN-1999; 99WO-US12697.
XX
XX 09-JUN-1998; 98US-0088805.
XX
XX (USDA) US DEPT OF AGRICULTURE.
XX
XX Gross KC, Smith DL;
XX

Db 904 tcaattattacatgtatcatgtgaggaacaaactttggcagcgtctggtgtccattta 963
Qy 980 ttgcaactagctacgatttatgtctctcgtcgatgaataatggtgtcgaatgaacaa 1039
Db 964 ttgctactagtattgactatgacacacttgatgaataatggtatttgagacaacaa 1023
Qy 1040 agtatggcacttgagacttacataaagaactatcaaatctatctgaacggcttagttt 1099
Db 1024 aatgggtcacctgaagatctgcatagagcaataaagctttgtgaacagcttttagtct 1083
Qy 1100 catcatatgctgcggtgactagcttggaagttaatacaagagcgcactatttatagataa 1159
Db 1084 ctggagatccagcgtgtgacagcacttggacaccagcagagggcccatcttttttagtgga 1143
Qy 1160 aatctggagcttgctgcttttttccaaactatgactctagatattcagtaaaagtcga 1219
Db 1144 aggctggtcttgctgctattcctgttaactacgacccaactcttttgcactgctg 1203
Qy 1220 cctttcagaataggccatacaatctgctcccatggtccatcagcaattcttcccgactgca 1279
Db 1204 catttgcaaacaggcattacaaacttgccaccatggtccaatcagcattcttcccgactgca 1263
Qy 1280 aaactgcogetttaacaaactcacaggttaactctcaagctcgagcaataaagatgaagc 1339
Db 1264 agaacactgtatttaatacacagcaggtacggtgctcaagtgctcagatgaagtgaactc 1323
Qy 1340 ctgcaggtggtgattctcttggcagtcatacaatgaagaaacgcctactgctgatgaca 1399
Db 1324 cagtcagcagaggttgccctggcagtcattcaatgaagagacatcatcttatgaagaca 1383
Qy 1400 gcgatacacttacagctaacgactatgggaacagaaaaacgctcacagagattcatcag 1459
Db 1384 g---tagttttacagtttggtgctattggaacagataaatacaacaagagacgtgctg 1440
Qy 1460 actatctgtgtacatgacaatgtaaatatagcatctcaatgaaggtattctctaaagaagc 1519
Db 1441 attattgtgtattcaacagatgccaagattgattcgaagagaaaagtttttgagagcg 1500
Qy 1520 gaaaggtaccttatctactgttatgtccgctggtcattgtctgttttctcgaatg 1579
Db 1501 gaaaatggccttgcttaacatgctcagctggcagtcattgcattgttttggatg 1560
Qy 1580 gaaactatcaggaactgtttatggtacattggataatccaaaaacttacatcagtgga 1639
Db 1561 gtcaattagcgaagactgcatggaagtttagaaaaaccgaaactaactttcagtaaa 1620
Qy 1640 acgfgaagttaagagctggtattacaagatttctctcagtggtttccgtgtgtctcc 1699
Db 1621 ccgtaaactcgagagcaggtgttaacaagatttctcactgagcattgctgtggccttc 1680
Qy 1700 cgaactgtggcgtgcatcatgatacatggaatgagcaggtgtctaggtccagtcagtgga 1759
Db 1681 cgaatatcgcccaattttgagacatggaatgctggtgtcttggccagctcactaa 1740
Qy 1760 gcggtctcaatgaaggtgtcaagaacttggcgaaacagagaataatggtcttacaaggttgctc 1819
Db 1741 ctggctgtgacgaggggaaagagatttaacatggtgcagagaaatggtttcacaaggttgctc 1800
Qy 1820 tgaagcgcaactgttaagctttcactccttaagtggagttcttctgttgaatgggttc 1879
Db 1801 taaaaggagacccctgagttcttcaactcactcagtggttagcccatccgtggagtggtg 1860
Qy 1880 gaggttcaatggtctcaagcagccctgacttggtaacaggctacatttaacgcgc 1939
Db 1861 aagctcttttagtggcacagagcagccactcagttggtataagactacattcaatgctc 1920
Qy 1940 ctggaggaatgatccactagcttttagacatggcaagtattgggaaaaaggtcagatattgga 1999
Db 1921 cagatggaaatgaacctttgcttttagatgaataccatgggcaaaaggtcaagtatgga 1980
Qy 2000 taaatggtgaagcgtaagtgcgaattggcctggatatacatagcacaaagcgactgcagca 2059
Db 1981 taaatggtcagagcctcgagccactggcctgcatataataatcatctggaaggttgatg 2040

Qy 2060 aatgcagttatgtggaacgttcaacgagagaagatgcggaactaactcggaacacctt 2119
Db 2041 tctgtactatactgctggtgttgatgagaaaaagtgcctaaactaactgtggtgagggct 2100
Qy 2120 ctocgagatggtacacatgttccacgactggtgctgaaacccaagtggaaacttgttagtag 2179
Db 2101 cacaagatggttacacgtaccccggtcttggtgtatcctactggaatttgttagttg 2160
Qy 2180 taticgaagaatgggaggttaatccaacagagaatttctctagtcagagatcaagataaa 2239
Db 2161 tattcggagaatgggagagatccttatgaatcaccttagtcaaaagagaatagggga 2220
Qy 2240 gaactcgaaaagtataaactgttccagtaactatggtctgaattcgc 2287
Db 2221 gtgtttgtgcgatataatgatgagtggaaccacagatttgaattggc 2268

RESULT 5

AAQ89602
ID AAQ89602 standard; cDNA: 2945 BP.
XX
AC AAQ89602;
XX
DT 21-NOV-1995 (first entry)
XX
Tomato beta galactanase coding sequence.
DE
XX
KW Beta galactanase; galactosidase; fruit; ripening; tomato;
firmness; texture; viscosity; genetic engineering; ss.
XX
OS Lycopersicon esculentum.

PN W09510622-A.

XX 20-APR-1995.

XX 10-OCT-1994; 94WO-GB02203.

XX 12-OCT-1993; 93GB-0020930.

XX (ZENE) ZENECA LTD.

XX Bird CR, Carey AT, Holt KA, Picard S, Schuch WW;

PI Seymour GB, Tucker GA;

XX WPI: 1995-161807/21.

DR Modifying cell wall metabolism in plants with a DNA construct -
that affects galactosidase activity, esp. for control of fruit
ripening, also transformed plant cells and derived plants

PS Claim 3; Page 30-31; 39pp; English.

XX Cell wall metabolism in plants can be modified by transforming the
plants with a DNA construct encoding a tomato galactanase which is
able to modify the activity of at least one galactosidase. Plants
transformed in this way have latered galactanase gene expression
CC which alters fruit-ripening characteristics. Reduction of gene
expression will increase fruit firmness, resistance to mechanical
damage, improve texture and resistance to disease and also increase
CC the viscosity of processed fruit. Increasing galactosidase activity
CC will also modify fruit texture and processing properties. This is
CC the tomato galactanase enzyme coding sequence.

SQ Sequence 2945 BP; 851 A; 525 C; 599 G; 870 T; 0 other;

Query Match 45.7%; Score 1168.2; DB 16; Length 2945;

Best Local Similarity 70.3%; Pred. No. 2.3e-226;
Matches 1579; Conservative 0; Mismatches 663; Indels 3; Gaps 1;

Qy 43 tttttttgaatgtgaaaaaatgctaaggagactaatgtgtgtgtattattgtt 102

Db 1422 ctgacaaattacacacgacactcttttctaaagtggaatttgggaacatgcattataac 1481
QY 1243 ctgctccatggtccatcagcattctccgactcaaaactgcggttttacacactgca 1302
Db 1482 ttgcacacctggtctatcagcattctccgactcaaaactgcggttttacacactgca 1541
QY 1303 caggttaactctcaagctcgagcataaagatgacgctgcaggtgggtggtctcttgg 1362
Db 1542 aggggttgctcaaaagtgcagatgaagatgactccagtcagtgaggtctctcatgg 1601
QY 1363 cagtcatacaatgaagaacgctactgctgatgacgagcatacacttacagctaacga 1422
Db 1602 gagtcatcaatgaagacgagcagcatgcgatgaaga--cgacactttcacagttgttgg 1658
QY 1423 ctatgggaacagaaaagctcaagagattcatcagactctgtgtgacatgacaaa 1482
Db 1659 ttattggagcagattaaatacaagagatgtatctgtactctgtgtatatgactgac 1718
QY 1483 gtaatatagcatctaatgaaggatttctaaagacggaaggatcccttatctactgtt 1542
Db 1719 attgagattgatccacagaaggatttttgaatagtgaattggccttggtactgtc 1778
QY 1543 atgtccgctggtcgtctgttgcattgttttcgtaattggaagaaactatcaggaaacttttat 1602
Db 1779 ttttctgctggcctgcattgcatgtattcgtgaatgttcaattagcaggaactgtgtac 1838
QY 1603 ggtacattggtataatccaaaacttacatcagtggtgcaacggtgaagttaagagctggtatt 1662
Db 1839 ggaagttagaanaacccaaaacttaacttccagcaacggtataaactcagagctggtgtg 1898
QY 1663 aacaagattctctcagtggttccctgtgtgtccgcaacggttggcgtgcattatgat 1722
Db 1899 aacaagattctctcagtggttccctgtgtgtccgcaacggttggcgtgcattatgat 1958
QY 1723 acatggaatcagaggtctcaggtccagtcacgttgagcgttctcaatgaagggtcaaga 1782
Db 1959 acatggaatcaggtgtgttctggaccagtttcaacttaattggaagggaacaaga 2018
QY 1783 aacttggtgaacagaaaatggtctacaaaggttggtctgaaggcgaaatcgtaagtctt 1842
Db 2019 gatttaacatgacagaaaatggtctacaaaggttggtctgaaggcgaaatcgtaagtctt 2078
QY 1843 cactccttaagtggagttctctgtgaatgggttcgaggttcaactaatgggtcacaag 1902
Db 2079 cattcactcagtggtgagccatccgtgagtggtggaaggctctttagtggtcagaag 2138
QY 1903 cagccctgactgtgtacaaaggtcatttaacggtcgtgaggaatgataccactagct 1962
Db 2139 cagccactcagttgtgtataaagactacatcattgctccagatggaaatgaaccttggct 2198
QY 1963 ttgacatgggcaagtatggaaaggtcagatattgataaaatgggtgaaggcgtaggtcgc 2022
Db 2199 ttgatatgataccatgggcaaaaggtcgaagtatggaataatggccagcctcgagcgc 2258
QY 2023 cattggcctggatacatagcaagcgtcagcagcaaatgcagttatgctgggaacgttc 2082
Db 2259 cactggcctgataataatcattctggaagtgtagtgtctgtaactataactggcgtgtt 2318
QY 2083 aacgagaagaagtcgggaactaactcgggaacacccctctcagagatggtacacatgtcca 2142
Db 2319 gatgagaaaagtgcactaactcgtgtgagggctcacaagaatggtacacagtcacc 2378
QY 2143 cgtatcgtggctgaaccagtggaactgttagtagtattcgaagaatggggaggtaat 2202
Db 2379 cgggttctggtgtactcctactggaataattgttagtattcagaggaatggggagagat 2438
QY 2203 ccaacaggaattctctagtcaggagcatcaagataaaagcaactcgaagaaactgttt 2262
Db 2439 ccttatggaatcattttagtcaaaagagaaataggaggtgttggctgatatatgatgag 2498
QY 2263 cagtaactatggtcgtgaattcgc 2287

Db 2499 tggcaaccacagttattgaattggc 2523
RESULT 7
AAC49769
ID AAC49769 standard; DNA; 2469 BP.
AC AAC49769;
XX 18-OCT-2000 (first entry)
XX Arabidopsis thaliana DNA fragment SEQ ID NO: 62367.
DE Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
OS Arabidopsis thaliana.
XX
XX EP1033405-A2.
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126284.
PR 29-MAR-1999; 99US-0126785.
PR 06-APR-1999; 99US-0127462.
PR 08-APR-1999; 99US-0128234.
PR 16-APR-1999; 99US-0128714.
PR 19-APR-1999; 99US-0129845.
PR 21-APR-1999; 99US-0130077.
PR 23-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132484.
PR 06-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 18-MAY-1999; 99US-0134370.
PR 19-MAY-1999; 99US-0134768.
PR 20-MAY-1999; 99US-0134941.
PR 21-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.

[illegible]

Db	1226	cagaggtttcaacactgcgaagggtcaatgcaccaaagcgtacacaggaataatggttcta	1282
Qy	1343	cagggtggattctcttgccagtcatacaatgaagaacgcctactgctgatcacagcg	1402
Db	1286	ctggtaacaaattctctgggatcatccaacgaagcgactcctcacagcgaatgaagctg	1345
Qy	1403	ataccttacagctaacggaactatgggaacagaaaaacgctcacaaagagattcatcagact	1462
Db	1346	gtacgtttgcaagaacagggccttggaacacaaatgaagcatgacttgggacaaatctgact	1405
Qy	1463	atctggtgatacagacaaatgtaatatagcatctaaatgaaggagattctcaagaacggaa	1522
Db	1406	attttgggtataaaacagacattacaatcggttccoggtgagacatttttgaagactgggt	1465
Qy	1523	aggatccttatcactgttatgtccgtggtcatgctctgcattgttttcgtcaatggaa	1582
Db	1466	attcaccctcttaacagttatgtcagctggacatgctctcatgcttgttgcgaatgctc	1525
Qy	1583	aactatcaggaaactgtttatggtacattggataatccaaaaacttacatcagttggcaacg	1642
Db	1526	agcttcaggaaactgcttatgaggagacttgaccaccacaaaactaacctttagccagaaga	1585
Qy	1643	tgaagttaagagctggtattaaacagagattctctctgctcaggttttcogttggtctccga	1702
Db	1586	tcaactacatgcagggtgcacaagagattgctctctgagtggttcgcagttgggtctccga	1645
Qy	1703	acgttggtgctgcattatgatcatggaatgcaggagttcttaggtccagtcacgtttgagcg	1762
Db	1646	acgttggatacaatttgaaacagtggaataaaggagttcttggctcggtcacattgaagg	1705
Qy	1763	gtctcaatgaagggttcagaaaacttggcgaaacagaaatggtcttcaaaagtttgggtctga	1822
Db	1706	gagtaaacacaggaaactgggacatgtcgaaatggaatggaatggtcctaaagattggtgtaa	1765
Qy	1823	aaggcgaatgtaagtctctactccttaagtggaggtctctctgttgaatgggttcogag	1882
Db	1766	aggtgaggtcttgagctcttcataccacacagagaggttcoggtgtgagatggactcaag	1825
Qy	1883	gttcactaatggtcctaaagacgcccctgacttggtaacaggtcacatttaacgcgcctg	1942
Db	1826	gactatttggcgaggagcaccattaaccttggcacaagcttgcgtttgtctacaccag	1885
Qy	1943	gaggaaatgatccactagcttttagacatggcgaagtatgggaaaaaggtcagatatggataa	2002
Db	1886	caggaaatgaaccattggcttttagacatgaacacaatgggaaaaaggtcaagtttgataa	1945
Qy	2003	atggtgaagcgtagtgctgcatttggcttggatgatacatagcacaaggcgactgcagcaaat	2062
Db	1946	acgtgagaataattggacgttcattggcttgatataaagctcaaggagatcttggacgtt	2005
Qy	2063	gcagttatgctgaacgcttcaacgagaagaagtcgcgaactaacctgcgcacaacctcttc	2122
Db	2006	gtactacgctggaactttcgacgacaagaataatgcttaagtcaattgttgggtgaagcttca	2065
Qy	2123	agagatggtaccattgttccacgatcgttggctgaaacccaagtggaaaaacttcttagtagtat	2182
Db	2066	aagaatggtaccatgtacctcgtttcatggtctaaatgc---tcaaaaccttatgttgtgt	2122
Qy	2183	tcgaagaatggggagggtaatccaacaggaaattctctagtccagagatcaagataa	2238
Db	2123	tcgaagagttaggtgtgatctcctaacgaaatctcttggtttaaagaacatgaaa	2178
RESULT 8			
AAC49810			
ID	AAC49810 standard; DNA; 2571 BP.		
XX			
AC	AAC49810;		
XX			
XX			
DT	18-OCT-2000 (first entry)		
XX			
XX	Arabidopsis thaliana DNA fragment SEQ ID NO: 62521.		
DE			

PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 02-JUL-1999; 99US-0142154.
PR 06-JUL-1999; 99US-0142053.
PR 08-JUL-1999; 99US-0142390.
PR 09-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 19-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144634.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 23-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 03-AUG-1999; 99US-0146389.
PR 04-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.

Db	1585	tcttgggcccggcggcattgcattggaatgcagggaataatggattgtcatcggcaga	1744
Qy	1799	aatgggtcttcaaaaggttggtctgaaaggcgaatcgttaagtcttcaactccttaagtggga	1858
Db	1745	aatggacataccaggcgggacgtgaaagtgaaagctgaatctctgcatttcccaaccaaca	1804
Qy	1859	gtctctctgtgaatgggttcgaaggttcaactaatgggtcctaaaaag---cagcccttgactt	1915
Db	1805	ctccctctatgtgatggatgggtgcctcttaactgtacaaaagcctcagcctttgacat	1864
Qy	1916	ggtacaaggctacatttaacgcgcctggaggaaaatgataccactagctttagacatggcaa	1975
Db	1865	ggcaacaagactattttgtatgcgcgaaggaaaatgagccgtctgtttggacatggaaag	1924
Qy	1976	gtatgggaaaaaggtccagatattgataaaatgggtgaaagcgttaggtcgccattggcctggat	2035
Db	1925	gaatgggaaaaaggtcagatttgggtgaaacggttgagagcatcgggaagatactgtgacagcat	1984
Qy	2036	acatagcaaaaggcgactgcagcaaatcgattatgctgggaacgttccacgagaagaagt	2095
Db	1985	---ttgcgactggtgactgcagtcattgtagctacacgtggaacataaaaccgacaacaa	2041
Qy	2096	gcgggaactaactcggacaacattctcagagatggataccattgtttccacgatactgtgcctga	2155
Db	2042	gccaaaccggctggtgtcaacctacacaaaggtggataccattgtcggagagcttgggttaa	2101
Qy	2156	aaccaagtggaaactgttagttagtatcttcgaagaatggggaggttaattcccaacaggaaatt	2215
Db	2102	agccgggccaataactgttcggctcatcttcgagggagcttggcggaaaacccataactgtct	2161
Qy	2216	ctctagtccaggagatcaagataaagaactcg	2246
Db	2162	ctttggtcaaaagatcagtatcagagatttg	2192

RESULT 11
AAT15173
ID AAT15173 standard; cDNA: 2778 BP.

AAT15173;

DT 02-JUL-199

Nasturtium xyloglucan-specific beta-galactosidase cDNA:

xyloglucan-specific beta-galactosidase; nasturtium; gelling;
transgenic plant; ripening; ss.

OS *Tropaeolum majus*.

	Key	Location/Qualifiers
AA		
FF		
FT	CDS	23..2596
FT		/tag= a
FT	sig_peptide	23..121
FT		/tag= b
FT	mat_peptide	122..2593
FT		/tag= c

PN WO9607743-A1.

14-MAR-1996

XX 07-SEP-1995: 95WO-GB02098

YY
PR 09-SEP-1994. 94EP-0306635

XX
PA
/UNTIL
/UNTIL.EVER NY

PA (UNIL) UNILEVER PLC.
yy

PI Chengappa S, De Silva J, Hellyer SA, Reid JSG;
vv

'DR WPI; 1996-171617/17.

DR P-PSDB; AAR95287.

PT New isolated xylo-glucan-specific beta-galactosidase and related DNA
AA
PT
PT - used for altering the gelling properties of xylo-glucan or for
PT altering plant characteristics e.g ripening

PS Claim 7: Fig 1: 36pp: English.

A cDNA clone (AAT15173) codes for a nasturtium xyloglucan-specific beta-galactosidase (XSBG) (see AAR95287) that catalyses the hydrolysis of terminal galactose residues from polymeric xyloglucan. It was obtained from a cDNA library of germinating nasturtium seeds using primers (see AAT15175-82) designed on the basis of isolated peptides of XSBG. The cDNA can be incorporated into a vector for use in recombinant production of mature XSBG, or can be used to alter plant characteristics, e.g. growth, texture and ripening.

Sequence 2778 BP: 820 A: 519 C: 652 G: 787 T: 0 other: xx

Query Match	27.8%	Score 708.8;	DB 17;	Length 2778;
Best Local Similarity	59.9%;	Pred. No. 1.1e-133;		
Matches 1297;	Conservative	0;	Mismatches 822;	Indels 45;
Gaps				5;

Qy 80 t g t t g t g t t a r a g t t a t t g t g t t c c a g t g a a g c t a g t g t t t 139
| | | | | | | | | | | | | | | | | | | | |
Db 75 t t t g c t a t t c t g t t t c c a a t t t g t t g t g t t t t t t t c t c t g -- c a t c c a a t g t g a 131

Qy 140 cttatqatqacagaqctataatcataaatqqaaaaaaattcttatttctqattcaa 199

132 cttatgatcatcgttctctcatcattgatgcccaacggaggctactgaatttcagcttcta 191

0v 200 ttcatattccaagaagcactccacagatattgccctgattcttatcaaaaggctaaagatg 259

Db 192 ttcat taccct cgaag tgt cct ggaat gt gccc ggcct ttt caaa caac taaa aag 251

[illegible]

352 371

0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100

[illegible][illegible][illegible]

=====

Q1

[illegible]

QY 360 TCGAATCCCAAGGAGGACCCAAATATATGCCCCAGATACCAAAATGAGTATGGAACCAATAG 615

Db 552 ttgcatcacaaggagggtcccataatcttgtcccagggttgaaaatgaatatgagataccg 611

Qy 620 aatgggaaattgggtgctcctggttaagcttatcaaaaatgggcagctcaaatggctgtag 679

Db 612 aacagttttatggggcaggggggaagttatatgccatgtggcccaatatggccattt 671

Qy 680 gttgaaaactgggtgtcccatggatcatgtgtaagcaagaggatgctcctgatcctgtga 739

Db 672 ctcaaatataggtgtaccttggataatgtgccagcaatatgatgctcctgatcctgtga 731

Qy 740 ttgatacttgtaatggccttctactggaagggttccgtcctaataagccttacaaccta 799

Db 732 tcaatacttgtaattccttctactgtgacaagttcataccaactctccgaacaagccca 791

Qy 800 aaatgtggacagaagtatggactggctggtatatacgaaattcgggtgccaatccctcaaa 859

Db 792 aatgtgactgagaactgacctggtggttttaaaacttttgggctagagactcctata 851
QY 860 gaccagccgaagacattgcatcttcagttgcccaggtttgttcagaaacaaatggttcattct 919
Db 852 gaccagcggagatattgctgttgcgttctctgcttcttcccaaaagggtggcagctac 911
QY 920 tcaattactacatgtatcatgaggagaaacaaattttggccggaacatcatcagggcttttca 979
Db 912 aaactactacatgtatcatggtggaacaaattttggtcgtacatcagcgaccattta 971
QY 980 ttcaactactacatgtatgtatgctcctcctcgtatgaataatggttgcgtgaagaacaa 1039
Db 972 ttaccacaagtattgattgatgctgccaatcgtatgtagtgcgttttaccctaggtctccaa 1031
QY 1040 agtatgggaccttgagagacttcaataaaagcttatcaagcttatcgaaccgcttttagtt 1059
Db 1032 aatggggacaccttaagaacctcaataaaagcttatcaagcttatcgaaccgcttttagtt 1091
QY 1100 catcatatgctgctggtgacttagtcttggaaagtaatacaagaggtctatgtttatagatcaa 1159
Db 1092 accatgaatctatgaacatgtcacttggccattggttagagccgatgtttatacggata 1151
QY 1160 aatctggagctgtgctgtgtttttatccaaactatgactctagatatcctgtaaaagtaa 1219
Db 1152 attcaggagaaatgtgctgcatcttctgcacacatagatgataaagaagacaaagttgtac 1211
QY 1220 cctttcagaatagccatataatctgctccatggttccatcagctatcttcccgactgca 1279
Db 1212 agttcgaaatgtgtcttaacacttctcctgctggtcagttagcattctcccgactgca 1271
QY 1280 aaactgcgttttaacaactgcaggttaactctcaagctcgaagctcgaacataaaagatgacgc 1339
Db 1272 agaattagctataaacacgcgaaggtcggttctcagatgtctacagttcacaatgttcc 1331
QY 1340 ctgacggtggtgattctcttggcagtcatacaatgaagaacgcctactcgtga----- 1394
Db 1332 ctgagaaataaatacatcagtcgctcctatccgacaaagccgacacctcttaaatggg 1391
QY 1395 -----tgacagggatacacacttacgtaacggaactatg----- 1427
Db 1392 aagtggttagagaagcaggatttggggagaagccgactttgttaaaagtagtttg 1451
QY 1428 -ggaacagaaaaacgtcacagagattcatcagactactctgtgtacatgacaaatgtaa 1486
Db 1452 tggatcatactacatactacaaagatactactgactactatgtgtacacaaagtataa 1511
QY 1487 atatagcatataatgaaggatttctaaagaaacggaaggtactctatctcactgttatgt 1546
Db 1512 ctgtgtggaagatgaagatctcttgaagaagggaagtaacccaggttctcttattgagt 1571
QY 1547 cgcgtggtcatgtcttgcattgtttcgttcaatggaaactatcaggaactgtttattggtta 1606
Db 1572 caaagggtcagtcctctcaagcttctgttgatcaggaaactcgaagcagtgcatgtggga 1631
QY 1607 cattgataatccaaactacatacagtggaacgtggaagtttaagagctgggtattaaaca 1666
Db 1632 atggatacagcttctgctttacataaagaagcctatttctcgaagcggggaagaagc 1691
QY 1667 agatttctcgtcaggtttctcgttgggtctccgaaacgttggcgtgcatattatgatacat 1726
Db 1692 atatcaaccttaagaatgtgtcgtggtctacagaatgctggaccatcttaaga---at 1748
QY 1727 ggaatcaggagttcaggtccagtcacgttggcgtctcctcaatgaaggtcgaagaact 1786
Db 1749 ggcagagcagactacaagcgtcagttatgaagggtcacaacatgggacactagact 1808
QY 1787 tggcgaacagaaatggtcttacaaggttgggtctggaagcgcaatcggttaagcttcaact 1846
Db 1809 tgccttaataatggtcttacaagattgagattgcgaggtgaacacctgggtatacaca 1868
QY 1847 ccttaagtggaagttctctgttgtaatgggttggaggttgcactaatgtgtcctcaaaagcagc 1906

Db 1869 acccaggtagttttagtgggttaaatggtgacgtcgaactcgaagcaccgcaaacagcagc 1928
QY 1907 cctgacttggtacaaggtcacatttaacgcgcctgaggaataatgatccactagctttag 1966
Db 1929 cattgacatggtacaaggttgttggaccacccccaggggagtgaaccacatcggtctg 1988
QY 1967 acatggcaagtattgggaaaaaggtcagatatggataaaatggtgaaggcgtaggtcgccatt 2026
Db 1989 acatgcttcataatgggaaaaggtctgctggttggaatggagaagaaatcgaagatact 2048
QY 2027 ggcctgatacatagacaagggcagactgcagca-----aatgcagttatgctggaacgt 2080
Db 2049 ggcctagaaaaagttctgtacatgatacttgttcccaagaatgcgactacagagcaaat 2108
QY 2081 tcaacgagagaagaatgcggaactaaactgcggaacacctctcagagatggtaccatgttc 2140
Db 2109 tcttcccggaacaaatgcttacgggagtgtggagatccaacgcaaaagatggtaccatgttc 2168
QY 2141 caagctgctggctgaaacacaaagtggaacttggaaacttcttagtagtattcgaagaatgggaagta 2200
Db 2169 ctggttctgttcaaacctgctggaacgttttgggttatattcgaaggaatgggtgaa 2228
QY 2201 atcc 2204
Db 2229 accc 2232

RESULT 12
AAZ29344
ID AAZ29344 standard; cDNA; 2972 BP.
XX AC AAZ29344;
XX 29-FEB-2000 (first entry)
XX Tomato beta galactosidase-7 cDNA.
DE Tomato beta galactosidase-7; TBG-7; clone pZBG2-1-4; fruit ripening;
KW Rutgers tomato plant; pectin; biofilm; gene mapping; ds.
XX Lycopersicon esculentum.
XX Key Location/Qualifiers
FH CDS 104..2716
FT FT /*tag= a /product= "Tomato beta galactosidase-7"
FT sig_peptide 104..208
FT mat_peptide /*tag= b 209..2713
FT /*tag= c
XX WO964564-A1.
PN 16-DEC-1999.
XX 08-JUN-1999; 99WO-US12697.
XX 09-JUN-1998; 98US-0088805.
XX (USDA) US DEPT OF AGRICULTURE.
XX Gross KC, Smith DL;
XX WPI: 2000-097532/08.
XX P-PSDB; AAY44309.
XX New beta-galactosidases, used to prepare transgenic plants with altered
PT fruit ripening -
XX Claim 2; Fig 2; 85pp; English.
XX The present sequence encodes tomato beta galactosidase-7 (TBG-7). This
CC is a cDNA insert of clone pZBG2-1-4 from a cDNA library derived from


```
Db 993 caactaaatggggccatttgaagatctgcatgctgtatataaaagctttgtgaaccagct 1052
Qy 1093 ttagtttcatcatat---gctgcygtgactagctcttgaagtaatacaagaggctcatgtt 1149
Db 1053 cttgtgtgctgctgactcaactcagtatataaactgggaccaaacagaggacacatgtc 1112
Qy 1150 tatagatacaaaatct-----gagcttgtgtct 1176
Db 1113 tctcgtggaacatccaacaacattggccaatatatgtcttaaatgaagggtatgcca 1172
Qy 1177 gcttttttaccactatgactctagatattcagtaaaagtcacacttttcagaataaggcca 1236
Db 1173 gcattttatgcaaatattgatgaacatagaatcagcaacagtgaaattttacggtcaagag 1232
Qy 1237 tacaatcgtccatcctggtgtccatcagcattctcccgactgcaaaactgcogtttacaac 1296
Db 1233 ttcaactttacctccatggf--cagtgtgtattctgccagattgcagaataacacagctttcaac 1291
Qy 1297 actgcacaggttaactctcaaaagctcgagcataaaagatgaagcctgcaggtggtgattc 1356
Db 1292 acagctaagggtggggccaaaacttcaatacaaaacagtgggctcagattctgttcaatt 1351
Qy 1357 tcttggcagtcatacaat----- 1374
Db 1352 ggggaataattcttgtttctacaagtattactaaaagcaagctcggaagttttttca 1411
Qy 1375 -----gaagaaacccctactgctgtagcacagcgatcacacttacaagctaa 1418
Db 1412 atcttggtagacattgaaggccacttgggtgtgggtgacaagaatttcaacttcaa 1471
Qy 1419 cggactatggaacagaaaacgtcacaaagagattcatcacactatctgtgtacatgac 1478
Db 1472 aggaatactgagcactgtaagtgcacaaaagaccagctctgtacacctgtgtgtatctgac 1531
Qy 1479 aaatgaaatagcatcatcattgaaggattttctaaaagcaacgaaaggatccttactcac 1538
Db 1532 caggataataattctgtatgacatcctcattttgggaggaataatgatgttagtccaac 1591
Qy 1539 tttatgtccgctggtgcattgtctgtcattgttttgcataatggaacactatcagaaactgt 1598
Db 1592 aatt-----gatattgtagcgtgctgattttgtgtcgcattttttaaaggcgagct 1645
Qy 1599 ttatggtacattgataataccaaaacttacatacagtggtgcaacgtgaagttaagagctgg 1658
Db 1646 tgcaggttagtgaaggcaaatggatggaagtggttcaaacctgttaagctggttcagg 1705
Qy 1659 tattaacaagaattctctgcagtggttccgttgggtcccccgaacgttggcgtgcatta 1718
Db 1706 atacaacacatactgctattatctgagacggtgggattgcagaattatgtgtgccttct 1765
Qy 1719 tgatactggaatcaggagttctaggtccagtcacgttggcggttccaatgaagggtc 1778
Db 1766 ggagaaggatggggcaggttttaaggctcagataaaagcttcaaggatgcaaaagcgggga 1825
Qy 1779 aagaaactggcgaacagaataatggtcttacaaggttggtctgaaggcgaatcgttaag 1838
Db 1826 tataactcacaacattttatggacctcacaggtggggcttggagcggaattcctcggga 1895
Qy 1839 tctcactccttaagtgggagttcttcttgaatgggttcaggttccactaatgggtcca 1898
Db 1886 agtatatgatgtcaatagtaactgaaagtcaggtgaggtgactgagtttcccactggtacaac 1945
Qy 1899 aagcagccccctgacttggtgacaaggtcacattttaacgcgcctggaggaatgatccaact 1958
Db 1946 tccgtcagctcttctcgttggtgacagaaaagtgtgctccacagcgcggaacagatccagt 2005
Qy 1959 agctttgacacatgcgaagtatggaaaaggtcagatatggaataatggtgaagcgctagg 2018
Db 2006 tgccttgatttttagtagcagtggaagggccaggcatgggttaattggccaccatgtagg 2065
Qy 2019 tcgcatttggcctggatcacatagca---caaggcgactgcagcaaatgcagttatgctgg 2075
Db 2066 aagatatggactttggttgaccacaaataatggtggaagaactgtgtattatcgtgg 2125
```

```
Qy 2076 aacgttcaacagagaagaagtcggaactaactcgcggaacaacctctcagagatgtacca 2135
Db 2126 tgcctaccactctgataaatgtaggacaaactgtgagagattactcagcctggtacca 2185
Qy 2136 tgttccacagatcgtgctgaaaccaaagtggaacttgttagtagttcgaagaatgggg 2195
Db 2186 tatactagatcgtgctgaaagacattaaataatgtactgtattcttgaagaacaga 2245
Qy 2196 aggttaatccaacaggaatttctctagtcagagatcaagataaagaactcg 2246
Db 2246 taaaactcgttgcatttccatttctacgcgttctactgaaccatttg 2296
```

RESULT 14

AAZ29343
ID AAZ29343 standard; cDNA; 749 BP.

XX AAZ29343;

AC AAZ29343;

DT 29-FEB-2000 (first entry)

XX Tomato beta galactosidase-6 cDNA.

XX Tomato beta galactosidase-6; TBG-6; clone pZBG2-1-4; fruit ripening;

XX Rutgers tomato plant; pectin; biofilm; gene mapping; ds.

XX Lycopersicon esculentum.

PH Key Location/Qualifiers

FT mat_peptide 1..747

FT /tag= a

FT /product= "Tomato beta galactosidase-6"

FT /note= "no stop codon given"

XX WO9964564-A1.

XX 16-DEC-1999.

XX 08-JUN-1999; 99WO-US12697.

XX 09-JUN-1998; 98US-0088805.

XX (USDA) US DEPT OF AGRICULTURE.

XX Gross KC, Smith DL;

XX WPI: 2000-097532/08.

XX P-PSDB; AAY44308.

XX New beta-galactosidases, used to prepare transgenic plants with altered

XX fruit ripening

XX Claim 2; Fig 2; 85pp; English.

XX The present sequence encodes tomato beta galactosidase-6 (TBG-6). This
is a cDNA insert of clone pZBG2-1-4 from a cDNA library derived from
breaker, turning and pink fruit pericarp from 'Rutgers' tomato plants.
CC TBG cDNA has been expressed in E.coli strain XLBlue MR. This is used
for modifying cell wall metabolism and controlling ripening of fruit by
altering activity of beta galactosidase II protein. Pectin with reduced
galactosyl content is produced for use in biofilms or solutions.
CC Fragments of the cDNA are used for gene mapping and for detecting
expression of beta galactosidase in plant tissue. Transgenic plants
with altered fruit ripening are produced by introducing DNA constructs
comprising a TBG cDNA associated with promoter/enhancer elements.

XX Sequence 749 BP; 205 A; 147 C; 186 G; 211 T; 0 other;

Query Match

Best Local Similarity 15.2%; Score 388.4; DB 21; Length 749;

Matches 537; Conservative 0; Mismatches 211; Indels 2; Gaps 2;

Qy	274	attgaaacttatgttttctggaatggacatggagccttctctctggaataataatttgaa	333	XX	25-FEB-2000;	2000EP-0301439.
Db	1	atccagacatatgtttttggaatgttcacatggccttctctctggaataataatttgaa	60	PF		
Qy	334	ggaagatatgatctgttagattcaataaattgacaaagagacagactttatgtcaat	393	PR	25-FEB-1999;	99US-0121825.
Db	61	ggaagatatgacctggtgaggtttgtataaacgattcagaagcaggcgtgtatgctcat	120	PR	05-MAR-1999;	99US-0123180.
Qy	394	ttacgtattggcccttaccgtctgctggaatggaactttgggggacattccctgtttgcta	453	PR	09-MAR-1999;	99US-0125788.
Db	121	cttcgaattggcccttaccgtttgtcagagtggaattttggagggtttccagtgatgctg	180	PR	25-MAR-1999;	99US-0126264.
Qy	454	aaatatgtcgtggttagaattagacaacaaatcagcctttttaagggtggcttatgcaa	513	PR	29-MAR-1999;	99US-0126785.
Db	181	aagtatgaactggcattagcttcagagctgataatgaaccttcaagaacgcaatgaaa	240	PR	06-APR-1999;	99US-0127462.
Qy	514	ggatttttcagaaaaatagtcacacatgatgaagtcagaaaaatttgtttgaaatctcaaggga	573	PR	08-APR-1999;	99US-0128234.
Db	241	gggtatgctgagaaaaattgttaacttgatgaag-atcataatcttttcgagctcaggg	299	PR	16-APR-1999;	99US-0128714.
Qy	574	ggaccaataattatggcccatgacaaaatgagtagtgaccagtagaatgggaaattggt	633	PR	19-APR-1999;	99US-0130077.
Db	300	ggtcccaataactctcacagattgagaatgagtagtggtggcctcaagccaagggtacttgg	359	PR	21-APR-1999;	99US-0130449.
Qy	634	gctcctgttaaagcttatacaaaaatggcagctcaaatggcgttagg-tttgaaaaactgg	692	PR	23-APR-1999;	99US-0130510.
Db	360	gcacgggacatcagtatccaacatggctgcaaatatggcagttgatttgaacacagg	419	PR	28-APR-1999;	99US-0130891.
Qy	693	tgctccatgatcatgtgtaagaagaaggatgctcctgatcctctgattgatacttggtaa	752	PR	30-APR-1999;	99US-0131449.
Db	420	cgtcccatgggtgatgtgcaagggaagaagatgcaccagatcctgtgatcaacacatgcaa	479	PR	04-MAY-1999;	99US-0132484.
Qy	753	tggcttactctggaagggttcgctcctaataagccttacaacctaataaattgtgacaga	812	PR	05-MAY-1999;	99US-0132485.
Db	480	tggtttctacttgataatttctcccaacaaacacccaataacactgcaatttggactga	539	PR	06-MAY-1999;	99US-0132486.
Qy	813	agtatgactggctgtgtatcgaataatcggtgtggtcccaattcctcaagagaccgcgaaga	872	PR	07-MAY-1999;	99US-0132487.
Db	540	agcttggagtggaatggttctcgaaatttggcggtcccccttcacagagaccagttcagga	599	PR	11-MAY-1999;	99US-0132863.
Qy	873	cattgcaattttcagttgccaggtttgttcagaaacaatggttcattctctcaattactacat	932	PR	14-MAY-1999;	99US-0134218.
Db	600	tttggaacttctgtgtcccaaatattacaaagaggaggtcttttgttaactattacat	659	PR	14-MAY-1999;	99US-0134219.
Qy	933	gtacatggagggaacaaattttgcccggacatcatcagggcttttctcattgcaactagcta	992	PR	14-MAY-1999;	99US-0134221.
Db	660	gtaccatgggggacgaactttggacgcactgcggtggggccattcatcaccagcta	719	PR	18-MAY-1999;	99US-0134370.
Qy	993	cgattatgctcctctcgatgaatatgg 1022		PR	20-MAY-1999;	99US-0134376.
Db	720	tgattatgctccccctcgacgagtgg 749		PR	21-MAY-1999;	99US-0134941.
RESULT	15			PR	21-MAY-1999;	99US-0135124.
AAC40074				PR	24-MAY-1999;	99US-0135253.
ID	AAC40074	standard; DNA; 1969 BP.		PR	25-MAY-1999;	99US-0136021.
XX				PR	27-MAY-1999;	99US-0136392.
AC	AAC40074;			PR	28-MAY-1999;	99US-0136782.
XX				PR	01-JUN-1999;	99US-0137222.
DT	17-OCT-2000	(first entry)		PR	03-JUN-1999;	99US-0137528.
XX				PR	04-JUN-1999;	99US-0137502.
DE				PR	07-JUN-1999;	99US-0137724.
DE				PR	08-JUN-1999;	99US-0138094.
XX				PR	10-JUN-1999;	99US-0138540.
KW				PR	10-JUN-1999;	99US-0138847.
KW				PR	14-JUN-1999;	99US-0139119.
KW				PR	16-JUN-1999;	99US-0139452.
XX				PR	17-JUN-1999;	99US-0139453.
OS				PR	18-JUN-1999;	99US-0139454.
XX				PR	18-JUN-1999;	99US-0139455.
XX				PR	18-JUN-1999;	99US-0139456.
XX				PR	18-JUN-1999;	99US-0139457.
XX				PR	18-JUN-1999;	99US-0139458.
XX				PR	18-JUN-1999;	99US-0139459.
XX				PR	18-JUN-1999;	99US-0139460.
XX				PR	18-JUN-1999;	99US-0139461.
XX				PR	18-JUN-1999;	99US-0139462.
XX				PR	18-JUN-1999;	99US-0139463.
XX				PR	18-JUN-1999;	99US-0139750.
XX				PR	18-JUN-1999;	99US-0139763.
XX				PR	21-JUN-1999;	99US-0139817.
XX				PR	22-JUN-1999;	99US-0139899.
XX				PR	23-JUN-1999;	99US-0140353.
XX				PR	23-JUN-1999;	99US-0140354.
XX				PR	23-JUN-1999;	99US-0140695.
XX				PR	28-JUN-1999;	99US-0140823.
XX				PR	29-JUN-1999;	99US-0140991.
XX				PR	30-JUN-1999;	99US-0141287.
XX				PR	01-JUL-1999;	99US-0141842.
XX				PR	01-JUL-1999;	99US-0142154.
XX				PR	06-JUL-1999;	99US-0142055.
PD	06-SEP-2000.			PR	06-JUL-1999;	99US-0142390.

Arabidopsis thaliana DNA fragment SEQ ID NO: 26952.
Hybridisation assay; genetic mapping; gene expression control;
protein identification; signal transduction pathway;
metabolic pathway; promoter; termination sequence; ss.

Arabidopsis thaliana.

EP1033405-A2.

06-SEP-2000.

Qy 2027 ggctggatacatagcacaggcgactgcagcaaatgcagttatgtctggaacgttcaacg 2086
||||| ||| ||| ||| |||||| || |||||| ||| ||| |||||| ||| |
Db 483 ggctgcataataagcaggttggttcttgacgcaggtgttcttatactggaacattcagag 542
||||| ||| ||| ||| |||||| || |||||| ||| ||| |||||| ||| |
Qy 2087 agaagaagtgcggaactaaactcgcggaacaaccttctcagagatggtaccatgttccacgat 2146
|| |||||| || |||||| ||| ||| |||||| |||||| |||||| ||| |||
Db 543 aggaagaagtcttaagaaacttggaagggttctcacaagatggtaccatgtcccaaggt 602
||||| ||| ||| ||| |||||| || |||||| ||| ||| |||||| ||| |
Qy 2147 cgtgctgaaaccaagtggaacttgtagtagtatttcgaagaatgggaggtaatccaa 2206
||||| |||||| |||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |
Db 603 cgtggctaaaaccaagtggcaatctattggttcttcttgaggagtgggagagaccoga 662
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |
Qy 2207 caggaatttctctagtcaggaga 2229
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |
Db 663 acggaatctcgttgtcagaaga 685
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |

Search completed: June 25, 2002, 06:29:18
Job time: 10523 sec

THIS PAGE BLANK (USPTO)

Db	11	ILLFSCI---FSAASASVSDHKAIIINGQKRILLISGSIHYPRSTPEMWPDLIOKADG	67
Qy	67	GLDVIETVYFWNGHSPGKYNEGRIYDLVRFKWQVAGLVNLRIGPYVCAEWNFGGF	126
Db	68	GLDVIQTVYFWNGHSPGKYNEGRIYDLVRFKWQVAGLVNLRIGPYVCAEWNFGGF	127
Qy	127	PWLKIVYVPGMEFTNPNQPKVAMQGFVQKIVNMKSENLFESQGGPITMAQIENYGPVE	186
Db	128	PWLKIVYVPGMEFTNPNQPKVAMQGFVQKIVNMKSENLFESQGGPITMAQIENYGPVE	187
Qy	187	WEIGAPGRAYTKAAQMAVGLTGVPMWCKQEDADPDVIDTCNCFYCEGFRPNKPYKPK	246
Db	188	WEIGAPGRAYTKAAQMAVGLTGVPMWCKQEDADPDVIDTCNCFYCEGFRPNKPYKPK	247
Qy	247	MWTEVMTGWTKFGGPIFORPAEDIAFVAREVQNGSFFNYMHGTTNFGRTSSGLFI	306
Db	248	MWTEVMTGWTKFGGPIFORPAEDIAFVAREVQNGSFFNYMHGTTNFGRTSSGLFI	307
Qy	307	ATSYDYDAPLDYGLLNEPKYGHLDLHKAIKLSEPALVSSVAATVSLGSNQEAHVYRSK	366
Db	308	ATSYDYDAPLDYGLLNEPKYGHLDLHKAIKLSEPALVSSVAATVSLGSNQEAHVYRSK	367
Qy	367	SGACAAFLSNYSRYSVKVTQFNRPYNLPWPSISILPDCKTAVYNTAQVNSOSSIKWTP	426
Db	368	SD-CAAEFLANYDAKYSVKSFGGGQYDLPWPSISILPDCKTAVYNTAQVNSOSSIKWTP	426
Qy	427	AGGSLWSQSYNETPADDSDTLTANGWEQKNVTRDSSDYLWYMTNVNIASNEGFLKNG	486
Db	427	VHSQFPWQSFTEETSSDETDTLLDGLYEQINIRDTDYLWYMTDITGSDEAFKNG	486
Qy	487	KDPYLTVMSAGHLVHFVNGKLSGTGYGTLDNPKITYSGNKLKRAKINKILLSVSVGLP	546
Db	487	KSPLLTIFSAGHALVNFINGQLSGVYVGSLENPKLFSQNVNLRSGINKLALLISVGLP	546
Qy	547	NGVHYDTWAGVLGVPVTLGSLNEGRNLAKQKWSYKVLKGESLSLSLSSSSSVVHR	606
Db	547	NGVHYDTWAGVLGVPVTLGSLNEGRNLAKQKWSYKVLKGESLSLSLSSSSSVVHR	606
Qy	607	GSLMAQKPLTWYKATFNAGNDPLALDMSMGKQIWIENGEGYRHWPGYIAGDCSK	666
Db	607	GSLMAQKPLTWYKATFNAGNDPLALDMSMGKQIWIENGEGYRHWPGYIAGDCSK	666
Qy	667	CSYAGTFNEKKQTCNGQPSQWYHVPWSLWPKPSNLLVWFEWCGNFTGISLYR	722
Db	667	CSYAGTYDDKRCRTHGEPSPQWYHVPWSLWPKPSNLLVWFEWCGNFTGISLYR	722
RESULT 2			
BDAL_LYCES	STANDARD:	PRT:	835 AA.
AC	P49980:		
DT	01-FEB-1996 (Rel. 33, Created)		
DT	01-FEB-1996 (Rel. 33, Last sequence update)		
DT	16-OCT-2001 (Rel. 40, Last annotation update)		
DE	Beta-galactosidase precursor (EC 3.2.1.23) (Lactase) (EXO-(1->4)-beta-D-galactanase).		
OS	Lycopersicon esculentum (Tomato).		
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;		
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;		
OC	Asteridae; euasterids I; Solanales; Solanaceae; Solanum.		
OX	NCBI_taxid=4081;		
[1]			
RP	SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.		
RC	STRAIN=CV. AILSA CRAIG; TISSUE=Pericarp;		
RX	MEDLINE=95357407; PubMed=7630937;		
RA	Carey A.T., Holt K., Picard S., Wilde R., Tucker G.A., Bird C.R.,		
RA	Schuch W., Seymour G.B.;		
RT	"Tomato exo-(1->4)-beta-D-galactanase. Isolation, changes during		
RT	ripening in normal and mutant tomato fruit, and characterization of a		
RL	related cDNA clone."		
RL	Plant Physiol. 108:1099-1107(1995).		
CC	-1- FUNCTION: INVOLVED IN CELL WALL DEGRADATION. DEGRADES		
CC	POLYSACCHARIDES CONTAINING BETA-(1->4)-LINKED GALACTANS, ACTING		

CC	AS AN EXO (1->4)-BETA-D-GALACTANASE. HAS A PH OPTIMUM OF 4.5.		
CC	-1- CATALYTIC ACTIVITY: Hydrolysis of terminal, non-reducing beta-D-		
CC	galactose residues in beta-D-galactans.		
CC	-1- SIMILARITY: BELONGS TO FAMILY 35 OF GLYCOSYL HYDROLASES.		
CC	-1- SIMILARITY: CONTAINS 1 SUEL-TYPE LECTIN DOMAIN.		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -		
CC	the European Bioinformatics Institute. There are no restrictions on its		
CC	use by non-profit institutions as long as its content is in no way		
CC	modified and this statement is not removed. Usage by and for commercial		
CC	entities requires a license agreement (see http://www.isb-sib.ch/announce/		
CC	or send an email to license@isb-sib.ch).		
CC	-----		
CC	EMBL: X83854; CAA58734.1; -		
DR	InterPro: IPR000922; Gal_Lectin.		
DR	InterPro: IPR001944; Glyco_hydro_35.		
DR	Pfam: PF02140; Gal_Lectin; 1.		
DR	Pfam: PF01301; Glyco_hydro_35; 1.		
DR	PRINTS: PR00742; GLHYDRLASE35.		
DR	ProDom: PD005612; Gal_Lectin; 1.		
DR	PROSITE: PS01182; GLYCOSYL_HYDROL_F35; 1.		
DR	PROSITE: PS02028; SUEL_LECTIN; 1.		
KW	Hydrolase; Glycosidase; Signal.		
FT	SIGNAL 1 22		
FT	CHAIN 23 835 BETA-GALACTOSIDASE.		
FT	DOMAIN 749 835 SUEL-TYPE LECTIN.		
FT	ACT_SITE 180 180 PROTON DONOR (POTENTIAL).		
FT	ACT_SITE 249 249 NUCLEOPHILE (POTENTIAL).		
SQ	SEQUENCE 835 AA; 93336 MW; 94C9685F95C4A646 CRC64;		
Query Match 74.5%; Score 2922.5; DB 1; Length 835;			
Best Local Similarity 72.2%; Pred. No 5,1e-196;			
Matches 518; Conservative 94; Mismatches 102; Indels 3; Gaps 2;			
Qy	6	VLLLVICLLDFFSSKVASVYDDRAIIINGKRKILISGSIHYPRSTPOMWPDLLIOKAKD	65
Db	7	MLLMLLLCL--WVSGIASVSDHKAIIIVNGQRKILISGSIHYPRSTPEMWPDLIOKAKE	64
Qy	66	GLDVIETVYFWNGHSPGKYNEGRIYDLVRFKWQVAGLVNLRIGPYVCAEWNFGG	125
Db	65	GGVDVIQTVYFWNGHSPGKYNEGRIYDLVRFKWQVAGLVNLRIGPYVCAEWNFGG	124
Qy	126	FPVWLKYVPGMEFTNPNQPKVAMQGFVQKIVNMKSENLFESQGGPITMAQIENYGPV	185
Db	125	FPVWLKYVPGMEFTNPNQPKVAMQGFVQKIVNMKSENLFESQGGPITMAQIENYGPV	184
Qy	186	EWEIGAPGRAYTKAAQMAVGLTGVPMWCKQEDADPDVIDTCNCFYCEGFRPNKPYK	245
Db	185	EWEIGAPGRAYTKAAQMAVGLTGVPMWCKQEDADPDVIDTCNCFYCEGFRPNKPYK	244
Qy	246	KWTEVMTGWTKFGGPIFORPAEDIAFVAREVQNGSFFNYMHGTTNFGRTSSGLF	305
Db	245	KWTEVMTGWTKFGGPIFORPAEDIAFVAREVQNGSFFNYMHGTTNFGRTSSGLF	304
Qy	306	IATSYDYDAPLDYGLLNEPKYGHLDLHKAIKLSEPALVSSVAATVSLGSNQEAHVYRS	365
Db	305	IATSYDYDAPLDYGLLNEPKYGHLDLHKAIKLSEPALVSSVAATVSLGSNQEAHVYRS	364
Qy	366	KSGCAAFSLNYSRYSVKVTQFNRPYNLPWPSISILPDCKTAVYNTAQVNSOSSIKMT	425
Db	365	ESGCAAFSLNYSRYSVKVTQFNRPYNLPWPSISILPDCKTAVYNTAQVNSOSSIKMT	424
Qy	426	PAGGSLWSQSYNETPADDSDTLTANGWEQKNVTRDSSDYLWYMTNVNIASNEGFLKN	485
Db	425	PVSRGFSWESNEDEAASHED-DFTFTVWGLLEQINIRDTDYLWYMTDIEDTPEGFLNS	483
Qy	486	GKDPYLTVMSAGHLVHFVNGKLSGTGYGTLDNPKITYSGNKLKRAKINKILLSVSVGL	545
Db	484	GNWPKLTVMSAGHLVHFVNGKLSGTGYGTLDNPKITYSGNKLKRAKINKILLSVSVGL	543
Qy	546	PNVGHYDTWAGVLGVPVTLGSLNEGRNLAKQKWSYKVLKGESLSLSLSSSSSVVHR	605


```

||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
544 PNVGPHFTWAGVLPVSLNGLNEGRDLTWQKWFYKVGKGLGALSLSHSGSPSVEW 603
QY RGLSMAOKQPLTWYKATFNAPGGNDPLALDMSKGGQIWIINGEGVGRHPWGYTAQGDPC 665
||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
604 EGSIVAOKQPLSWTKTTFNPDGNEPLALDMSKGGQIWIINGEGVGRHPWGYTAQGDPC 663
QY KCSYAGTFNEKKKOTNCQGPSQRYHVPWSLKPESGNLLVVEEWGNGPTGISLVR 722
||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
664 VCNVTGWFEDEKCLTNGEGSQRYHVPWSLKPESGNLLVVEEWGNGPTGISLVR 720

RESULT 3
ID BGAL_ASPOF STANDARD; PRT; 832 AA.
AC P45582;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DE Beta-galactosidase precursor (EC 3.2.1.23) (Lactase).
OS Asparagus officinalis (Garden asparagus).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Asparagales; Asparagaceae;
OC Asparagus.
OX NCBI_TaxID=4686;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. LIMBRAS 10; TISSUE=Sphear;
RX MEDLINE=95303968; PubMed=7784512;
RA King G.A., Davies K.M.;
RT "Cloning of a harvest-induced beta-galactosidase from tips of
RL Plant Physiol. 108:419-420(1995).
CC -1- CATALYTIC ACTIVITY: Hydrolysis of terminal, non-reducing beta-D-
CC galactose residues in beta-D-galactosides.
CC -1- SIMILARITY: BELONGS TO FAMILY 35 OF GLYCOSYL HYDROLASES.
CC -1- SIMILARITY: CONTAINS 1 SUEL-TYPE LECTIN DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
EMBL: X77319; CAA54525.1;
DR InterPro; IPR000922; Gal_lectin.
DR InterPro; IPR001944; Glyco_hydro_35.
DR Pfam; PF02140; Gal_Lectin; 1.
DR Pfam; PF01301; Glyco_hydro_35; 1.
DR PRINTS; PR00742; GLHYDRLASE35.
DR ProDom; PD005612; Gal_lectin; 1.
DR PROSITE; PS01182; GLYCOSYL_HYDROL_F35; 1.
DR PROSITE; PS02228; SUEL_LECTIN; 1.
KW Hydrolase; Glycosidase; Signal.
FT SIGNAL 1 25 POTENTIAL.
FT CHAIN 26 832 BETA-GALACTOSIDASE.
FT DOMAIN 741 832 SUEL-TYPE LECTIN.
FT ACT_SITE 183 183 PROTON DONOR (POTENTIAL).
FT ACT_SITE 252 252 NUCLEOPHILE (POTENTIAL).
SQ SEQUENCE 832 AA; 92213 MW; 94ABDC61EC4164AE CRC64;

Query Match 73.08; Score 2863; DB 1; Length 832;
Best Local Similarity 71.48; Pred. No. 7.2e-192;
Matches 514; Conservative 96; Mismatches 102; Indels 8; Gaps 5;

QY 6 VLLLVICLIDFFS--SVKASVYDDRAIILGKRLKILSGSIHYPRSTPQWPDLIQKA 63
||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
6 VLMVALLAAVAPPAVTVYDHSVILGQRRILSGSIHYPRSTPQWPDLIQKA 65
QY 64 KDGGLDVIETVYVFWNGHEPSPGKYNFEGRYDLVRFKMWQVAGLYVNLRIQGYVCAEWNF 123

```

```

||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
66 KDGGLDVIETVYVFWNGHEPSPGKYNFEGRYDLVRFKMWQVAGLYVNLRIQGYVCAEWNF 125
QY GGFVWLKLYVPGMEFRTNNQPKFVAMOGFVOKIVNMKSENLFSOGGPIITMAOIENEYG 183
||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
126 GGFVWLKLYVPGMEFRTNNQPKFVAMOGFVOKIVNMKSENLFSOGGPIITMAOIENEYG 185
QY PVEWEIGAPGKAYTKWAAQMAVGLKTVGPWIMCKQEDAPDVIDTCNGFYCEGFRPNKPY 243
||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
186 PVEYDGAAGKSYTNWAAKMAVGLKTVGPWIMCKQEDAPDVIDTCNGFYCEGFRPNKPY 245
QY KPMWTEVMTGWYTKFQGPPIQRPDAEDIAFSVARFVQNGSFFNYMTHGTFNRTSSG 303
||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
246 KPMWTEVMTGWYTKFQGPPIQRPDAEDIAFSVARFVQNGSFFNYMTHGTFNRTSSG 305
QY LFIATSYDYDAPLDEYGLLNEPKVGHRLDLHKAIKLSPALVSSVAAVTSLSGNOEAHVY 363
||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
306 PFISTSYDYDAPLDEYGLLNEPKVGHRLDLHKAIKLSPALVSSVAAVTSLSGNOEAHVY 365
QY RSKGCAACAAFLNSYDSRYSVKVTQNPYPNLPWMSISILPDKCTAVYNTAQVNSQSSIK 423
||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
366 RSKS--SCAAFLANFSRYIATVTENGHYNLPWMSVSLPDKCTAVYNTARVGAQTTMK 424
QY MTPAGGGLSWOSYNEETPTADSDTLFANGLWEOKNVTDRSDSYLWYNTVNNVNASGFL 483
||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
425 MQYL--GGFSWKAYTETDLDND--NTFTKDLGVEQLSTTWDRSDYLYTYTYVDIARNEEFL 482
QY KNGKDPYLYTVMSAGHVLHVFNGKLSGTVCTYGLONPKLTSYGNVKLRAGINKISLSYSV 543
||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
483 KTGKPYLYTVMSAGHVLHVFNGKLSGTVCTYGLONPKLTSYGNVKLRAGINKISLSYSV 542
QY GLPNVGVHYDTWAGVLPVTLGSLNGEGRNLAKQWYKVLGKAGESLSLSHSGSSSVE 603
||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
543 GLPNVGNHFEFTWNTGLGVTGLTGLNEGRDLSLQKWIYQIHLGHEITLSHSLTGSSNVE 602
QY WVRGSLMAOKQPLTWYKATFNAPGGNDPLALDMSKGGQIWIINGEGVGRHPWGYTAQGD 663
||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
603 WGEAS--OKQPLTWYKATFNAPGGNDPLALDMSKGGQIWIINGEGVGRHPWGYTAQGD 659
QY CSKSYAGTFNEKKKOTNCQGPSQRYHVPWSLKPESGNLLVVEEWGNGPTGISLVR 723
||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
660 CGSCDYRGTYNEKKKLSNCGEASQRYHVPWSLKPESGNLLVVEEWGNGPTGISLVR 719

RESULT 4
ID BGAL_DIACA STANDARD; PRT; 731 AA.
AC Q00662;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Putative beta-galactosidase precursor (EC 3.2.1.23) (Lactase)
DE (SRL2 protein).
DE CARSL2.
OS Dianthus caryophyllus (Carnation) (Clove pink).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Caryophyllales; Caryophyllaceae; Dianthus.
OX NCBI_TaxID=3570;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. WHITE SIM; TISSUE=Petal;
RX MEDLINE=91329738; PubMed=1868223;
RA Raghothama K.G., Lawton K.A., Goldsbrough P.B., Woodson W.R.;
RT "Characterization of an ethylene-regulated flower senescence-related
RL gene from carnation."
CC -1- CATALYTIC ACTIVITY: Hydrolysis of terminal, non-reducing beta-D-
CC galactose residues in beta-D-galactosides.
CC -1- TISSUE SPECIFICITY: SENESCING FLOWER PETALS.
CC -1- INDUCTION: BY ETHYLENE.
CC -1- SIMILARITY: BELONGS TO FAMILY 35 OF GLYCOSYL HYDROLASES.
CC -----

```


Qy 181 EYGPVWEIGAPKAYTKWAAQAVGLKTPWIMCKOEDAPDPVIDTCNGFYCBGFRPN 240
Db 182 EYGNVSSYGAEGRAKIDWCANMNSLDIGVPMICQPHAPQPMIETCNGFYCDQYKPS 242
Qy 241 KPYPKMTEWNTGTYTFKFGPIQORPAEDAFSAVAREVQNGSFNFYMYHGGTNGFRT 300
Db 243 NPSPGKMTEWNTGTYTFKFGPIQORPAEDAFSAVAREVQNGSFNFYMYHGGTNGFRT 302
Qy 301 SSGFIATSYDADPLDEYGLNPKYGLHRLKAIKLSPPALVSSVAAVTSLSGNOEA 360
Db 303 AGGPVITTSYDADPLDEYGLNPKYGLHRLKAIKLSPPALVSSVAAVTSLSGNOEA 362
Qy 361 HVRKSGAACAAFLSNYSRYSKVTQFQNPYNLPWPPSISILPCKTAYNTAQVNSOSS 420
Db 363 TVY-STNEKSSCFIGNVATADALVNFKGDYNPVPAWSVLPDCKDEAYNTARVNTQTS 421
Qy 421 SIKM-TPAGGGLSQ-SYNEETPTADDSTLTANGLWEQKNVTRDSSDYLTWMTNVN 475
Db 422 IITEDSCDEPKLWTRPEFTQTKILKSGDLIAKGLVDKQVNDASDYLTWMTNRVH 481
Qy 476 IASNEGFLKNGKDPY-LTVMSAGHVLHVFVNGKLSGTGYTGLDNPKLTYSGNYKL 529
Db 482 L-----DKKDPWRSNMSLRVHSNAHLHAYVNGKYVGNQIVRDNKFDYRPEKKYNL 533
Qy 530 RAGINKISLLSVSGLPVNGVHYDTWNAVLGPVTLGSLNEG-SSRLAKQKWSYKVL 586
Db 534 VHGTHLALLSVSGLQNGYFPFESGPTGNGPVKLVGKGDTEKDKLSKHQWDYKIGL 593
Qy 587 KGESLSLHSL-SGSSSVWVRGSLMAQOKPLTWYKATFNAPGNDPLALDMSMGKQI 644
Db 594 NGFNHLKFSMSAGHHHRKSTELKPADRM-LSWYKANEKAPLGDGPVIVDLNGLGKGEV 652
Qy 645 WINGEGRHRWPGYTA--QGCSCSKSYAGTFNEKKQCNQOPSORVHVRPSWLKPSG- 701
Db 653 WINGQISGRYPFNSNSDEGCTEEDYRGEYGSCKAPWCCKPQRTQVHVRPSFLNDKXG 712
Qy 702 NLLVVFEEWGNPT 715
Db 713 NTITLFEEMGDPS 726

RESULT 6

ID BGAL_XANMN STANDARD; PRT; 598 AA.
AC P48982;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Beta-galactosidase precursor (EC 3.2.1.23) (Lactase).
GN BGA.
OS Xanthomonas manihotis.
OC Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
OC Xanthomonas.
OX NCBI_TaxID=43353;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 22-38.
RC STRAIN=7AS1;
RX MEDLINE=96121416; PubMed=8563148;
RA Taron C.H., Benner J.S., Hornstra L.J., Guthrie E.P.;
RT "A novel beta-galactosidase gene isolated from the bacterium
RT Xanthomonas manihotis exhibits strong homology to several eukaryotic
RT beta-galactosidases.";
RL Glycobiology 5:603-610(1995).
CC -1- FUNCTION: PREFERENTIALLY HYDROLYZES BETA(1->3) GALACTOSYL LINKAGES
CC -1- OVER BETA(1->4) LINKAGES. HAS A PH OPTIMUM OF 4.5.
CC -1- CATALYTIC ACTIVITY: Hydrolysis of terminal, non-reducing beta-D-
CC galactose residues in beta-D-galactosides.
CC -1- SIMILARITY: BELONGS TO FAMILY 35 OF GLYCOSYL HYDROLASES.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: L35444; AAC41485.1; -
DR InterPro: IPR001944; Glyco_hydro_35.
DR Pfam: PF01301; Glyco_hydro_35; 1.
DR PROSITE: PS01182; GLYCOSYL_HYDROL_F35; 1.
KW Hydrolase; Glycosidase; Signal.
FT SIGNAL 1 21
FT CHAIN 22 598 BETA-GALACTOSIDASE.
FT ACT_SITE 184 184 PROTON DONOR (POTENTIAL).
FT ACT_SITE 260 260 NUCLEOPHILE (POTENTIAL).
SQ SEQUENCE 598 AA; 66085 MW; DB4C3F05E01435BF CRC64;

Query Match 13.4%; Score 525.5; DB 1; Length 598;
Best Local Similarity 26.0%; Pred. No. 3.5e-29;
Matches 191; Conservative 90; Mismatches 279; Indels 175; Gaps 26;

Qy 1 MLRTNVLVLLVLCILDFESSVKASVSY-----DDRAIIINGKRKILISGSIHYPRSTPQM 55
Db 1 MLRTTLPALVLAALALPAAATPESWTFGTGCTQFVRDCKPYQLLSAIFQIPRAY 60
Qy 56 WPDLLQAKAGDGLDVIETVYFVWNGHEPSPGKYNFEGRYDLVRFIRKVRAGLYVNLRTGP 115
Db 61 WKDLQKARALGNTVETVYFVWNLVEPQGGQDFDSGNNDVAFAFYKEAAQGLVNLRTGPP 120
Qy 116 YVCAEWNNGGPPVWLKYVPNGHEFRTNQPFKVMQGFQKIVNMKSENLENSESQGGPIIM 175
Db 121 YACAEWAGGYPVWLFGKGNTRVRSRDRPFLAASQAYLDALAKQV--QPLLHNGGPIIA 178
Qy 176 AQIENEYGPVEWIEGAPGKAYTKWAAQAVGLKTPWIMCKOEDAPDPVIDTCNGFYCE 235
Db 179 VQVENEYSY-----ADHAY--MADNRAMYVKAQFGLKALLFTSGADML---ANGTLPD 228
Qy 236 -----GFRPN-----KPKPKMTEWNTGTYTFKFGPIQORPAEDAFSAVARE 278
Db 229 TLAVVNFAPGEAKSAFADKLINFRPDQPRMVGWYAGWFDHNGKP---HAATDARQAEEF 285
Qy 279 --VQNGSFFNYHYHGGTNGFRTSSGLF-----IATSYDADPLDEYGLNPK 326
Db 286 EMILRQGHSAFLYMFIGTSGFGFMNGANFQNPSPDHYAPQTSYDYDAILDEAG-HPTPK 344
Qy 327 YGHLRDLHKAIKLSEPALVSSVAAVTSLSGNOEAHVYRKSACAAFLSNYSRYSVKVT 386
Db 345 FALMED-----AIARVTGV----- 358
Qy 387 FQNRPNLP-PWISILPCKTAYNTAQVNSQSSIKMTAGGGLSWQSYNEETPTADD 445
Db 359 ---QPPALPAPITTTTL-----ATPLRESASL-----WD--NLPTPIAD 394
Qy 446 SDTLTANGLWEQKNVTRDSSDYLTWMTNVNITASNGFLKNGKDPVLTWMSAGHVLHVFVN 505
Db 395 T-----PQPMQFGQDYGYILYRTTITGPR-----KGP-LYLGVDVDRVARVVD 437
Qy 506 GLKSGTYGTLDNPKLTYSGNVKLRAGINKISLSVSVGLPNVGVHYDTWNAAGVLPVTL 565
Db 438 QRPVGSVERRLOOVSL-----EVEIPAGOHTLDLVENSGRINYGTRMADRAGLVDVPLL 493
Qy 566 SGLNFGSRLAKQKWSYKVLKGLGESLSLHLSGSSSVWVRGSLMAQOKPLTWYKATNA 625
Db 494 D-----SQOL-----TGQAFFLPNR-----TPDSIRGWTGKAVQGFAPFRTGLRI 534
Qy 626 FGNPDPLALDMSMGKQIWIWNGEGVGRHWPGYIAQGCSCSKSYAGTFNEKKQCNQOP 685
Db 535 GTPPTDY-LDMRAFCKGFANGVNLGRHW-----NTGPTQTYLALRPSRARV 580
Qy 686 SORVHVRPSWLKPS 700
Db 581 TTRWSSSTWMLHPS 595

RESULT 7
 BGAL_HUMAN STANDARD; PRT; 677 AA.
 AC P16278;
 DT 01-AUG-1990 (Rel. 15, Created)
 DT 01-AUG-1990 (Rel. 15, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE Beta-galactosidase precursor (BC 3.2.1.23) (Lactase) (Acid beta-galactosidase).
 GN GLB1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RC TISSUE=Testis;
 RX MEDLINE=90062209; PubMed=2511208;
 RA Morreau H., Galjart N.J., Gillemans N., Willemssen R., van der Horst G.T.J., D'Azzo A.;
 RT "Alternative splicing of beta-galactosidase mRNA generates the classic lysosomal enzyme and a beta-galactosidase-related protein."; J. Biol. Chem. 264:20655-20663(1989).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=90262647; PubMed=2111707;
 RA Yamamoto Y., Hake C.A., Martin B.M., Kretz K.A., Ahern-Rindell A.J., Naylor S.L., Mudd M., O'Brien J.S.;
 RT "Isolation, characterization, and mapping of a human acid beta-galactosidase cDNA."; DNA Cell Biol. 9:119-127(1990).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Placenta;
 RX MEDLINE=89061717; PubMed=3143362;
 RA Oshima A., Tsuji A., Negao Y., Sakuraba H., Suzuki Y.;
 RT "Cloning, sequencing, and expression of cDNA for human beta-galactosidase."; Biochem. Biophys. Res. Commun. 157:238-244(1988).
 RN [4]
 RP VARIANTS MORQUIO B LEU-273; HIS-482 AND CYS-509.
 RX MEDLINE=92026088; PubMed=1928092;
 RA Oshima A., Yoshida K., Shimmoto M., Fukuhara Y., Sakuraba H., Suzuki Y.;
 RT "Human beta-galactosidase gene mutations in morquio B disease."; Am. J. Hum. Genet. 49:1091-1093(1991).
 RN [5]
 RP VARIANTS GM1 CYS-49; THR-51 AND CYS-201.
 RX MEDLINE=91353572; PubMed=1909089;
 RA Nishimoto J., Nanba E., Inui K., Okada S., Suzuki K.;
 RT "GM1-gangliosidosis (genetic beta-galactosidase deficiency): identification of four mutations in different clinical phenotypes among Japanese patients."; Am. J. Hum. Genet. 49:566-574(1991).
 RN [6]
 RP VARIANTS GM1 THR-51; ARG-123; CYS-201; CYS-316 AND GLN-457.
 RX MEDLINE=91328151; PubMed=1907800;
 RA Yoshida K., Oshima A., Shimmoto M., Fukuhara Y., Sakuraba H., Yanagisawa N., Suzuki Y.;
 RT "Human beta-galactosidase gene mutations in GM1-gangliosidosis: a common mutation among Japanese adult/chronic cases."; Am. J. Hum. Genet. 49:435-442(1991).
 RN [7]
 RP VARIANT GM1 HIS-482.
 RX MEDLINE=93138608; PubMed=1487238;
 RA Mosna G., Fattore S., Tubiello G., Brocca S., Trubia M., Gianazza E., Gatti R., Danesino C., Minelli A., Piantanida M.;
 RT "A homozygous missense arginine to histidine substitution at position 482 of the beta-galactosidase in an Italian infantile GM1-gangliosidosis patient."; Hum. Genet. 90:247-250(1992).
 RN [8]
 RP VARIANTS GM1 CYS-208; ARG-578; HIS-590 AND GLY-632.

RX MEDLINE=94027054; PubMed=8213816;
 RA Boustany R.-M., Qian W.-H., Suzuki K.;
 RT "Mutations in acid beta-galactosidase cause GM1-gangliosidosis in American patients."; Am. J. Hum. Genet. 53:881-888(1993).
 RN [9]
 RP VARIANT GM1 MET-82.
 RX MEDLINE=94256487; PubMed=8198123;
 RA Chakraborty S., Rafi M.A., Wenger D.A.;
 RT "Mutations in the lysosomal beta-galactosidase gene that cause the adult form of GM1 gangliosidosis."; Am. J. Hum. Genet. 54:1004-1013(1994).
 RN [10]
 RP VARIANTS MORQUIO B HIS-83 AND CYS-482.
 RX MEDLINE=96049832; PubMed=7586649;
 RA Ishii N., Oohira T., Oshima A., Sakuraba H., Endo F., Matsuda I., Sukegawa K., Orii T., Suzuki Y.;
 RT "Clinical and molecular analysis of a Japanese boy with Morquio B disease."; Clin. Genet. 48:103-108(1995).
 RN [11]
 RP VARIANTS GM1 HIS-59; ASN-591 AND CYS-591.
 RA Morone A., Bardelli T., Donati M.A., Giorgi M., Di Rocco R., Gatti R., Taddeucci G., Ricci R., D'Azzo A., Zammarchi E.;
 RT "Identification of new mutations in six Italian patients affected by a variant form of infantile GM1-gangliosidosis with severe cardiomyopathy."; Am. J. Hum. Genet. 61:A258-A258(1997).
 RN [12]
 RP VARIANTS GM1 H-59; S-121; C-208; M-240 AND N-491, AND VARIANTS P-10; C-521 AND G-532.
 RX MEDLINE=99268417; PubMed=10338095;
 RA Silva C.M.D., Severini M.H., Sopelsa A., Coelho J.C., Zaha A., d'Azzo A., Giugliani R.;
 RT "Six novel beta-galactosidase gene mutations in Brazilian patients with GM1-gangliosidosis."; Hum. Mutat. 13:401-409(1999).
 CC -!- FUNCTION: CLEAVES BETA-LINKED TERMINAL GALACTOSYL RESIDUES FROM GALGIOSIDES, GLYCOPROTEINS, AND GLYCOSAMINOGLYCANS.
 CC -!- CATALYTIC ACTIVITY: Hydrolysis of terminal, non-reducing beta-D-galactose residues in beta-D-galactosides.
 CC -!- SUBCELLULAR LOCATION: Lysosomal.
 CC -!- ALTERNATIVE PRODUCTS: BETA-GALACTOSIDASE AND BETA GALACTOSIDASE-RELATED PROTEIN ARE PRODUCED BY ALTERNATIVE SPLICING OF THE SAME GENE.
 CC -!- DISEASE: DEFECTS IN GLB1 ARE THE CAUSE OF GM1-GANGLIOSIDOSIS, AN AUTOSOMAL RECESSIVE DISORDER WITH THREE MAJOR CLINICAL PHENOTYPES THAT ARE DISTINGUISHED ACCORDING TO THE AGE OF ONSET AND SEVERITY OF SYMPTOMS: INFANTILE, JUVENILE AND ADULT. THE INFANTILE FORM IS RAPIDLY PROGRESSIVE AND RESULTS IN SEVERE CENTRAL NERVOUS SYSTEM DEGENERATION AND VISCEROMEGALY, WITH DEATH USUALLY BETWEEN THE FIRST AND SECOND YEAR.
 CC -!- DISEASE: DEFECTS IN GLB1 ARE THE CAUSE OF MUCOPOLYSACCHARIDOSIS IV B (OR MORQUIO B SYNDROME) WHICH IS CHARACTERIZED BY SEVERE BONE DEFORMITIES WITHOUT CNS INVOLVEMENT.
 CC -!- SIMILARITY: BELONGS TO FAMILY 35 OF GLYCOSYL HYDROLASES.

 This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

 EMBL; M27507; AAA51819.1; -;
 EMBL; M34423; AAA51823.1; -;
 EMBL; M22590; AAA51822.1; -;
 PIR; A32611; A32611.
 PIR; A31673; A31673.
 PIR; A32688; A32688.
 PIR; B37086; B37086.
 MIM; 230500; -;

CC -!- SIMILARITY: BELONGS TO FAMILY 35 OF GLYCOSYL HYDROLASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF006749; AAB81350.1; -;
DR EMBL; AF029974; AAB86405.1; -;
DR InterPro; IPR001944; Glyco_hydro_35.
DR Pfam; PF01301; Glyco_hydro_35; 1.
DR PRINTS; PR00742; GLHYDRASE35.
DR PROSITE; PS01182; GLYCOSYL_HYDROL_F35; 1.
KW Hydrolase; Glycosidase; Lysosome; Signal; Glycoprotein.
FT SIGNAL 1 24
FT PROPEP 25 29 BY SIMILARITY.
FT CHAIN 30 669 BETA-GALACTOSIDASE.
FT ACT_SITE 189 189 PROTON DONOR (BY SIMILARITY).
FT ACT_SITE 269 269 NUCLEOPHILE (POTENTIAL).
FT CARBOHYD 27 27 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 248 248 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 465 465 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 499 499 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 547 547 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 557 557 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 483 483 R -> P (IN REF. 2).
SQ SEQUENCE 669 AA; 75229 MW; 35B84933BB5E2F76 CRC64;

Query Match 12.5%; Score 490.5; DB 1; Length 669;
Best Local Similarity 26.3%; Pred. No. 1.1e-26;
Matches 182; Conservative 91; Mismatches 271; Indels 147; Gaps 26;

QY 8 LLLVLCILDFSSVKAS-----VSDDRILINIKRILISGSIHYRPTQMPMDLIQ 61
DB 12 LLLVPLLLGARGLRNASORTKIDYGHNRFLKDGQPFYISGSIHYRPRFYWKDRLL 71
QY 62 KAKDGLDVIETVFWNGHEPSPGKYNEGKYDVLRFKMWORAGLYNLRIGPVYCAEW 121
DB 72 KMKMAGLNAIQTVFWNHEPQGGQYQSGEHDEYFLKLAHELGLLVLRPGPYICAEW 131
QY 122 NFGGPPVNLKYVPGFEFRINQPFKVMQGFQKIVNMKSNLENFESOGGPIIMAQIENE 181
DB 132 DMGGLPAWLLKESILRSSDPDLAANDKWLGVLLPKMKP--LLYQNGSPIITQVENE 189
QY 182 YGP-----VWEIGAPGRAYTKAAQ---MAVGLKTGVPMCKQEDAPD 223
DB 190 YGSYFTCDYDLRFLQRPRDLRGDVLFTTDDGAHEKFLQCGALQG---IVATVDFGPD 246
QY 224 PVIDTCNCFYCGFRPNKPYKPKMTETVWGTYTFEGGPIPORPAEDIAFVAREVQNG 283
DB 247 ANITA--AFQIQ--RKSEPRGLPVNSEFTYDGLDHWGQPHSRVREVASSL-HDVLAHG 301
QY 284 SFENYMYHGGNFGRFTSGGLPI-----ATSDYDAPLDYEGLLNEPKYGLHRLDHLKA-- 336
DB 302 ANVNLVMTFGTTFAYW-NGANIPYQOPTSYDYDAPLSEAGDLTD-KYFALROVIRKFE 359
QY 337 -----IKLSEPALVSSAAVTSLSNOBAHVYRSKGACAFLSNYSRYSKVYTFQNR 390
DB 360 KYPEGFIPPESTPKFAYKVALOKLTVEDA-----LNLVCPA--GPRIKSLYPLTIQVKQ 412
QY 391 PYNLPPWISILPDKCTAVYNTAQNVSQSSSIKMTMPAGGGLSQSYNEETPTADSDTLT 450
DB 413 YGFVLVYRTLPQDCS-----NFTPLUSSPLNGVR-----DRAVYAVDG---V 451
QY 451 ANGLWEQKNVTRDSSDYLYMTNVNVIASNEGFLKNGKDPYLFVMSAGHYLHVFNKGLSG 510
DB 452 PQGVLSR-----YVITLNTIGQAGATLD-----LLVENMGRVNY----GRVIN 491
QY 511 TVYGLDNPKLTYSGNVKLKRAGINIKILLSVSVGLPNNGVHYDTWNAGVGLVPTLSGLNE 570

DB 492 DFKGLISN--LTLGSSVLTDWMIFPLDTEDA-----VRSHLGWGH-----GRNH 533
QY 571 GSRNLAKQKWSYKVLKESLSLHSGSSSVWVRGSLMAQKQLPTWYKATFNAPGGND 630
DB 534 GRQD-----NKAFAHSSNYTLP-----AFYAGNFSIPSGIP 565
QY 631 PLALD-----MASMGRGQIWINGEVGRHWPG 657
DB 566 DLPQDTFFIQFSGWTKGQVWINGENLGRYWP 596

RESULT 9
BGAL_MOUSE
ID BGAL_MOUSE STANDARD; PRT; 647 AA.
AC P23780;
DT 01-NOV-1991 (Rel. 20, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DE 01-MAR-2002 (Rel. 41, Last annotation update)
DE Beta-galactosidase precursor (EC 3.2.1.23). (Lactase) (Acid beta-
DE galactosidase).
GN GLB1 OR GLB-1 OR BGL.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=91076843; PubMed=2124109;
RA Nanba E., Suzuki K.;
RT "Molecular cloning of mouse acid beta-galactosidase cDNA: sequence,
RT expression of catalytic activity and comparison with the human
RT enzyme."
RL Biochem. Biophys. Res. Commun. 173:141-148(1990).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=DBA/2J;
RX MEDLINE=91298941; PubMed=1906271;
RA Nanba E., Suzuki K.;
RT "Organization of the mouse acid beta-galactosidase gene."
RL Biochem. Biophys. Res. Commun. 178:158-164(1991).
CC -!- FUNCTION: CLAVES BETA-LINKED TERMINAL GALACTOSYL RESIDUES FROM
CC GANGLIOSIDES, GLYCOPROTEINS, AND GLYCOSAMINOGLYCAN.
CC -!- CATALYTIC ACTIVITY: Hydrolysis of terminal, non-reducing beta-D-
CC galactose residues in beta-D-galactosides.
CC -!- SUBCELLULAR LOCATION: Lysosomal.
CC -!- SIMILARITY: BELONGS TO FAMILY 35 OF GLYCOSYL HYDROLASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M57734; AAA37293.1; -;
DR EMBL; M75122; AAA37292.1; -;
DR EMBL; M75137; AAA37292.1; JOINED.
DR EMBL; M75107; AAA37292.1; JOINED.
DR EMBL; M75108; AAA37292.1; JOINED.
DR EMBL; M75109; AAA37292.1; JOINED.
DR EMBL; M75111; AAA37292.1; JOINED.
DR EMBL; M75112; AAA37292.1; JOINED.
DR EMBL; M75113; AAA37292.1; JOINED.
DR EMBL; M75114; AAA37292.1; JOINED.
DR EMBL; M75115; AAA37292.1; JOINED.
DR EMBL; M75116; AAA37292.1; JOINED.
DR EMBL; M75117; AAA37292.1; JOINED.
DR EMBL; M75118; AAA37292.1; JOINED.
DR EMBL; M75119; AAA37292.1; JOINED.
DR EMBL; M75120; AAA37292.1; JOINED.

```
DR EMBL: M75121; AAA37292.1; JOINED.
DR PIR: A37086; A37086.
DR MGI: 88151; G1b1.
DR InterPro: IPR001944; Glyco_hydro_35.
DR Pfam: PF01301; Glyco_hydro_35; 1.
DR PRINTS: PR00742; GLHYDRLASE35.
DR PROSITE: PS01182; GLYCOSYL_HYDROL_F35; 1.
KW Hydrolase; Glycosidase; Lysosome; Signal; Glycoprotein.
FT SIGNAL 1 24
FT PROPEP 25 29
FT CHAIN 30 647
FT ACT_SITE 189 189
FT ACT_SITE 269 269
FT CARBOHYD 27 27
FT CARBOHYD 248 248
FT CARBOHYD 500 500
FT CARBOHYD 504 504
FT CARBOHYD 510 510
FT CARBOHYD 544 544
FT CARBOHYD 557 557
FT CARBOHYD 617 617
FT CONFLICT 517 517
FT CONFLICT 539 539
FT SEQUENCE 647 AA; 73121 MW; 0E68EA66A10803A CRC64;

Query Match 12.2%; Score 479.5; DB 1; Length 647;
Best Local Similarity 26.1%; Pred. No. 6.2e-26;
Matches 176; Conservative 85; Mismatches 257; Indels 157; Gaps 23;

QY 25 VSDRAIIINGKRKILISISIHYPRTQWPDLIQAKDGGLDVETVYFNGHBPSP 84
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 35 LDYSRDLKDGDPFRYISGSIHFRIYFREDLLKMKMAGLNAQMYVWNFHEPOP 94
QY 85 GKNFEGRYDLVRFKIVQRAVNLRIQYVCAENFGSPVWLKAYVPCMEPTNNQP 144
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 95 GQFEGDRDVEHFQIAHELGLLVILRPGYICAEWMDGMLPAWLEKOSIVLRSDPD 154
QY 145 FKVAMQGVQKIVNMKSENLFESQGGPIIMAQIENEYGP-----VWEI 189
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 155 YLVAVDKWLAVLLPKMKP--LLYQNGGPIITVQVENEYGYFACDYDLRFVHFRYHL 212
QY 190 GARGKATKWAQMAVGLKTVGWIMCKQEDAPDPVIDTONGFYCEGFRNPKYKPK--- 246
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 213 GNDVILFTTDCASEKM-LKCGTLQDLVATVD-----FGTGNNI-TQAFLVQRKEPFGPL 265
QY 247 MWEVMTGWTKFGGPIQPRAEDIAFVARFVQNNGSFFNYMYHGGTNF---GRSS 302
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 266 INSEFYGLDHDGKPHSTVKTATSLYNLLA-RGANVLYNFIGGTNFAYWNGANTP 324
QY 303 GLFIATSYDYDAPLDEYGLLNEPKYGLHRLDHLKAIK-----LSEPALVSSVAAVTS 354
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 325 YEQPTSYDYDAPLSEAGDLTK-KYFALREVIOFMKEVPEGPIPPPTPKAYGKVALRKP 383
QY 355 GSNOEAIHVRSKSGACAAFLSNVDSRVKVFQNRPNLPWISITLPCKTAYVNTAQ 414
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 384 KTVAAELGILCPNGP-----VKSLYPLTFTQVKQYFGVLYRTLLPQDC----- 427
QY 415 VNSQSSIKMTAGGGLSWQSYNEETPTADDSLTLTANGLWEQKNVTRDSSDYLWYTNV 474
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 428 --SNPKIFSSPNG-----YVSV-----DGVPPQGLDRNLT-ALNIRKAGAT 474
QY 475 NIASNEGLKNGKDPYLTVMASGHLVHFVNGKLSGTVYGTLDNPKLITYSGNVKLKRA 534
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 446 -----YVSV-----DGVPPQGLDRNLT-ALNIRKAGAT 474
QY 535 KISLLSVSGVLPVNG-----VHYDTWAGVLGVPVTLISGLNEGRNLAKQKWSKV 585
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 475 -LDILVENMGRVNGYGRINDFKGLISNMTINSTVLTNTVFPPLTEAM-VRNHLWGREAS 532
QY 586 LKGESLSLHISGSSSVFVWVRSGLMAQKPLTWYKATFNAPGGNDPLALD---MASMG 641
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 533 DEGH-LDGRSTNSDDL-----ILP-----TFYVGNFSIPSGIPDLPDPTFIQPGWSK 580
```



```
Query Match      10.0%; Score 393; DB 1; Length 1006;
Best Local Similarity 23.1%; Pred. No. 1.2e-19;
Matches 198; Conservative 102; Mismatches 318; Indels 238; Gaps 38;

Qy 25 VSYDDRAIIINGKRIISGSIHPRSTP--QMPDLDLQKADGGLDVIYVFWNGHEP 82
Db 46 VTWDDKSLFINGERIMFSGBFH-PFRLPVKELQDIPQVKALGFCNCSFYVDWALVEG 104
Qy 83 SPGVNFEGRYDLVRFIKWVRAGLYVNLGRPGYVCAEMNFGPPVWLKYVPGMEFRNN 142
Db 105 KPGEVRADGIDLEPFFDAEAGIYLARCPYINASSGGPGWQVRVNG-TLRSSD 163
Qy 143 QPFKVMQGFQKIVNMKSNLFESQGPPIIMAIQIENIEYGEVWEIGAPKAYTKWAAQ 202
Db 164 KAYLDATDNYVSHAATIAKQI--TNGPILYQENIYTSKSGVGEFPPVIMQVYED 221
Qy 203 MAVGLKTCVPMI-----MCKQEDADP---VIDTCN-----GFYC-----EGF 237
Db 222 QARNAGVVIPLINDNASAGSNAPGTGKGAVDIYGHDSYPLGFCANPTVWPMSGDLPTNF 281
Qy 238 R-----PNKPKPKMWTVMWTGYTKFGGPIPORPAEDIAFVAR-FVONGSF-----FNY 288
Db 282 RTLHLQESPTTPYAIVERQGSYDPWGGPGFAACSELLNFEFVYKNDFSQIATMNL 341
Qy 289 YMYHGGTNFGRS--SGLFIATSYDYDAPLDEYGLLNEPKYGHLDLHKAIKLSEPALVS 346
Db 342 YMIFGGTNWNLGYPNGY---TSYDYGSAVTESRNITREKYSSELKLLGNFAKVPGLYLA 398
Qy 347 SYAAVTSIG--SNOEHAHYRKSACAAFL-----SNYDSRYSVKVTQFONRPNLPPHSIS 400
Db 399 SPGNLTTSYGADTTDLTVTLPLGNSTGFFVVRHSDYSSEESTSYKLR-----LPTSAGS 453
Qy 401 I-LPCKTAVNTAQVNSOSSSIKWTTP---AGGGL-----SWOSVNEETPTADSDT 448
Db 454 VTIQ-----LGTTLNCRDSKIHTDHNVSCTNIIYSTAEVFTWKE-----ADGKVL 503
Qy 449 LTANGLME-----QKNVT-----461
Db 504 VLYGGAGHEHAIATSKSNVTVIEGSEGISKQTSVVVGVWDVSTPRRIIQVGDILKIL 563
Qy 462 ---RDSYLYWYMNVTIASNEGFLKNGKDPYLTVMAGHVLHVFNVGKLSGTVYGTLDN 518
Db 564 LLDRNSAYNVVWPQIATDGTSPGSTPEKVASIIVKAGYLVR-----TAY-----609
Qy 519 PKLYSGNVKLKRLAGINKISLSVSVGLPNVGVHYDTWAGVLPVTLISLNGSGSRNLAKQ 578
Db 610 --LKSG-LYLTADFNATTSVEV-IGVPSTAK-----NLFINGDKTSHVDKNG-----654
Qy 579 KWSYKVLKGLGESLSLHLSLG-----SSSVE---WVRGSLMAQKQPL-----616
Db 655 IWSATVDYNAPDISLPSLKLDWKYVDLTPETQSSYDSDSLWPAADLKQTKNLTSLTPT 714
Qy 617 -----TWYKATFNAGNDPLALDM--ASMGKQIWIINGEGVGRHPGYIAQ 661
Db 715 SLYSSDYGFHTGYLLYRGHFTATGNSTFAIDTQGSAGFAGSVWNLNGTLYLG-SWTGLYAN 773
Qy 662 GDCSKCSYAGTFNEKKCO-----TNGQGPQSQ 687
Db 774 SD-----YNATYNLPOLQAGKTYITVVDNMGLEENWTVGEDLMKSPRGISTSLPDGQ 828
Qy 688 RWHVHPSWLKPSGNL 703
Db 829 A---APISW-KLTGNL 840

RESULT 11
ID BGAM_HUMAN STANDARD; PRT; 546 AA.
AC P16279;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Beta-galactosidase-related protein precursor.
```

```
GN GLB1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., AND PLACENTAL, PARTIAL SEQUENCE.
RC TISSUE=Testis;
RX MEDLINE=90062209; Pubmed=2511208;
RA Morreau H., Gallart N.J., Gillemans N., Willemsen R.,
RA van der Horst G.F.J., D'Azzo A.;
RT "Alternative splicing of beta-galactosidase mRNA generates the
RT classic lysosomal enzyme and a beta-galactosidase-related protein.";
RL J. Biol. Chem. 264:20655-20663(1989).
CC -!- FUNCTION: THIS PROTEIN HAS NO CATALYTIC ACTIVITY.
CC -!- ALTERNATIVE PRODUCTS: BETA-GALACTOSIDASE AND BETA GALACTOSIDASE-
CC RELATED PROTEIN ARE PRODUCED BY ALTERNATIVE SPLICING OF THE SAME
CC GENE.
CC -!- SIMILARITY: BELONGS TO FAMILY 35 OF GLYCOSYL HYDROLASES.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M27508; AAA35599.1; -.
DR PIR; B32688; B32688.
DR MIN; 230500; -.
DR InterPro; IPR001944; Glyco_hydro_35.
DR Pfam; PF01301; Glyco_hydro_35; 1.
KW Signal; Alternative splicing; Glycoprotein.
FT SIGNAL 1 23
FT CHAIN 24 546 BETA-GALACTOSIDASE-RELATED PROTEIN.
FT CARBOHYD 26 26 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 116 116 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 333 333 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 367 367 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 411 411 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 414 414 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 424 424 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 546 AA; 60551 MW; 2B1A73EDAF9E966C CRC64;

Query Match      5.8%; Score 227.5; DB 1; Length 546;
Best Local Similarity 20.4%; Pred. No. 1.8e-08;
Matches 150; Conservative 77; Mismatches 226; Indels 283; Gaps 29;

Qy 6 VLLLVICLLDFSSVKAS-----VSYYDDRAIIINGKRIISGSIHPRSTPQMPDLD 59
Db 9 LLLLVILLGLPTGLRNATORMFEIDYSRDSFLKDGQFPRIYSGSIHYSRVPREYWDK 68
Qy 60 IQAKADGGLDVIETYVFWNGHEPSPGKYNFEGRYDLVRFIKWVRAGLYVNLGRPGYVCA 119
Db 69 LLKMKMAGLNAIQT-----82
Qy 120 EWNFGGFPVWLKYVPGMEFRNNQPFKVMQGFVQKIVNMKSNLFESQGPPIIMAIQIE 179
Db 83 -----LPG-----SCGVVGVSPSAQ 97
Qy 180 NEYGPVWEIGAPGKAYTKWAAQMAVGLKTVPTWIMCKQEDADPDPVIDTCNGFYCEGFRP 239
Db 98 DEASPL-----SEWRA-----SYNAGSNITD---AFLSQ--RK 126
Qy 240 NKPKPKMWTVMWTGYTKFGGPIQRPAEDIAFVARFVQNGSGFFNYMYHGGTNF-- 297
Db 127 CEPKPLINSEFYTGWLDHWGPHSTIKTEAVASSLYDILA-RGASVNLVWFIGTNPAY 185
Qy 298 --GRTSSGLFATSYDYDAPLDEYGLLNEPKYGHLD-LHKAIKLSEPALVSS-----YA 349
Db 186 WNGANSPYAAQPTSYDYDAPLSEAGDLTE-KYFALRNIIQKFEKVPGEPIPPSTPKFAYG 244
```



```
Db 225 -----DD-----AYSDKSMKVTVAFNQFG-----PNCQRMPRARYG 257
Qy 550 -VH-----YDTWAGVLG-----PVTSLNGEGRNLAQKQWYKVLKGESLSLHSLSGS 599
Db 258 LVHVANNNDPWTIATGSSNPILS-----EGNSTAPNE-SYK---KQVIRIGCKTSS 310
Qy 600 SSVIEW-----RGSMAQKQPLTWYKATFNAPGNN 629
Db 311 SCNNVWQSTQDFEYNGAYFVSSGYEGGNIYTKKE-----AFNVENG 354

RESULT 13
Y321_MYCGE
ID Y321_MYCGE STANDARD; PRT; 934 AA.
AC P47563;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical lipoprotein MG321 precursor.
GN MG321.
OS Mycoplasma genitalium.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
OC Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2097;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 33530 / G-37;
RX MEDLINE=96026346; PubMed=7569993;
RA Fraser C.M., Gocayne J.D., White O., Adams M.D., Clayton R.A.,
RA Fleischmann R.D., Bult C.J., Kerlavage A.R., Sutton G., Kelley J.M.,
RA Fritchman J.L., Weidman J.F., Small K.V., Sandusky M., Fuhrmann J.L.,
RA Nguyen D.T., Utterback T.R., Saudek D.M., Phillips C.A., Merrick J.M.,
RA Tomb J.F., Dougherty B.A., Bott K.F., Hu P.-C., Lucier T.S.,
RA Peterson S.N., Smith H.O., Hutchison C.A. III, Venter J.C.,
RA "The minimal gene complement of Mycoplasma genitalium."
RL Science 270:397-403(1995).
CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a lipid anchor
(Potential).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U39713; AAC71543.1; -
CC TIGR; MG321; -
CC PROSITE; PS00013; PROKAR_LIPOPROTEIN; 1.
CC Hypothetical protein; Lipoprotein; Membrane; Signal;
CC Complete proteome.
FT SIGNAL 1 24 POTENTIAL.
FT CHAIN 25 934 HYPOTHETICAL LIPOPROTEIN MG321.
FT LIPID 25 25 N-ACYL DIGLYCERIDE (POTENTIAL).
SQ SEQUENCE 934 AA; 102999 MW; EDAA036543736C6F1 CRC64;

Query Match 3.0%; Score 118.5; DB 1; Length 934;
Best Local Similarity 21.2%; Pred. No. 1.5;
Matches 149; Conservative 90; Mismatches 257; Indels 207; Gaps 38;

Qy 135 GMEPRTNQPKVAMQGVQKIVNMKSENLFESQGGPIIMAIQIENEYGPVWEIGAPG 193
Db 185 GREVKQNNQPKVLSSKDFERGFEYIILSSNLGFNRNGYFIDLMLGLD-----VEKTVG--- 236
Qy 194 KAYTKWAAQMAVGLKGVPMIMCKQEDAPD-----PVITDTCNGFCYCEGFR---PNK----- 241
Db 237 -----MDKNTGSSDGNKGKGIETDENVDENYRSYDDNKNFVYL 275
Qy 242 ----PYKPKMWTETWTKYFGGPIQRPADIAF-----SVARFVQNG-----SFF 286
Db 276 TSPFPFLSMMSK-----EFTFPPIHTHPKVALKLGDSPLUKNNNNRKILDQANT 328
```

```
Qy 287 NY-YMYHGGTNFGRT---SSGLFIATSYD-----YDAPLDEYGLLNEPKYGHRLD 332
Db 329 NFDGIYGGVNAWKDTWSVGPYYVESFNQAGIVFKRNOIYDAITP-----NLPK--TRQE 382
Qy 333 LHKAIKLSEPALVS-----SYAAVTSLSNQBAHVYRSKSGACAAFLSNVDSRY- 381
Db 383 NEKPI-----PAIVSYFQPGATPEVFSYIAGLSASAVPYSQOQDARSRENGTDLRLW 438
Qy 382 -----SVKVTFCNRPYNLPPWSISILPCKTAVYNTAOV-----NSOSSS 421
Db 439 KIQTAAQAVTYSCKPY-----VANDSTVQLNANITETEAFLYNSESEE 484
Qy 422 IKMTPAG--GGLSQSYN-EETPTADSDTLT-----ANGLMEQKNVTRDSSDYL 468
Db 485 ALTIRAGINGLINWKNLAIDLPSNGDVNYSTVPFGIFKEKPANGTSGGTNTDGIENDY- 543
Qy 469 WYMTNVN-----IASNEGFLKNGKDPYLTVMASGHV-LHVFVNGKLSGTVYGTLDNPKLT 522
Db 544 YKINNQRGLIPEQTGTFQDKN-----VLDTATVLSYSSTKTNGAQVTRAST----- 595
Qy 523 YSGNVKLKAGINKISLSVSV-----GLPNVG-----VHYDTWAGVLGPVTLISLN 569
Db 596 -SGS-----SSQTSQVSSKQSVTKQSFISALKKVGFTGNPLHFN-----KLGNASLSSNQ 647
Qy 570 EGSRLAKQKWSYKVLKGLKSLSLHSL-----SGSSSVVWVRGSLMAQKQPLTWYKATFN 624
Db 648 VDIYNALKQALTELGNNGENLIPEILGDAQGTNRNWIIG--LSVLGFSWSPDYD 705
Qy 625 APGNDPLALDMSMGKQIWIINGEGVGR-----HWPYIAGDCSKCSYAGTFN--EKKC 678
Db 706 GVG-----TWLDAATQ-----LNSEGIGEVITYNSGSHIVRTLLAASQNNVFNQIENKL 755
Qy 679 QTNCGQPSQRWYHVP RS--WLKPSGNLLVVFEEWGNPTGISL 719
Db 756 QNNTTTNGKMCWCITSAIDLKDDPYVIKNGFTNGNGTSASL 798

RESULT 14
FLT3_HUMAN
ID FLT3_HUMAN STANDARD; PRT; 993 AA.
AC P36888; Q13414;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DE FL cytokine receptor precursor (EC 2.7.1.112) (Tyrosine-protein kinase
DE receptor FLT3) (Stem cell tyrosine kinase 1) (STK-1) (CD135 antigen).
GN FLT3 OR STK1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94111906; PubMed=7507245;
RA Small D., Levenstein M., Kim E., Carow C., Amin S., Rockwell P.,
RA Witte L., Burrow C., Ratajczak M.Z., Gewirtz A.M., Civin C.I.;
RA "STK-1, the human homolog of Flk-2/Flt-3, is selectively expressed in
RA CD34+ human bone marrow cells and is involved in the proliferation of
RA early progenitor/stem cells."
RL Proc. Natl. Acad. Sci. U.S.A. 91:459-463(1994).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=93357464; PubMed=8394751;
RA Rosnet O., Schiff C., Pebusque M.J., Marchetto S., Tonnellet C.,
RA Toiron Y., Birg F., Birnbaum D.;
RA "Human FLT3/FLK2 gene: cDNA cloning and expression in hematopoietic
RA cells."
RL Blood 82:1110-1119(1993).
RN [3]
RP SEQUENCE OF 783-942 FROM N.A.
RC TISSUE-Testis;
RX MEDLINE=91169547; PubMed=2004790;
```


RT 1 in the t(8;13)(p11;q12) myeloproliferative syndrome.";

RL Blood 92:1735-1742(1998).

RN [3]

RX MEDLINE=99107818; PubMed=9889006;

RA Kulkarni S., Reiter A.J., Smedley D., Goldman J.M., Cross N.C.P.;

RT "The genomic structure of ZNF198 and location of breakpoints in the

RT t(8;13) myeloproliferative syndrome.";

RL Genomics 55:118-121(1999).

RN [4]

RX MEDLINE=99107818; PubMed=9889006;

RA Blakely S., Wall M.;

RT Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.

RN [5]

RX MEDLINE=98361795; PubMed=9694738;

RA Still I.H., Cowell J.K.;

RT "The t(8;13) atypical myeloproliferative disorder: further analysis of

RT the ZNF198 gene and lack of evidence for multiple genes disrupted on

RT chromosome 13.";

RL Blood 92:1456-1458(1998).

RN [6]

RX MEDLINE=98167848; PubMed=9499416;

RA Smedley D., Hamoudi R., Clark J., Warren W., Abdul-Rauf M., Somers G.,

RT Venter D., Fagan K., Cooper C., Shipley J.;

RT "The t(8;13)(p11;q12) rearrangement associated with an atypical

RT myeloproliferative disorder fuses the fibroblast growth factor

RT receptor 1 gene to a novel gene RAMP.";

RL Hum. Mol. Genet. 7:637-642(1998).

RN [7]

RX MEDLINE=98085877; PubMed=9425908;

RA Xiao S., Nalabolu S.R., Aster J.C., Ma J., Abruzzo L., Jaffe E.S.,

RT Stone R., Weissman S.M., Hudson T.J., Fletcher J.A.;

RT "FCFRL is fused with a novel zinc-finger gene, ZNF198, in the t(8;13)

RT leukaemia/lymphoma syndrome.";

RL Nat. Genet. 18:84-87(1998).

CC -!- FUNCTION: May function as a transcription factor.

CC -!- SUBCELLULAR LOCATION: Nuclear (Potential).

CC -!- DISEASE: Involved in a t(8;13)(p12;q12) chromosomal translocation

CC which involves FCFRL AND ZNF198. The resulting transcript is a

CC possible candidate for stem cell leukemia lymphoma syndrome/SCLL.

CC -!- SIMILARITY: CONTAINS 5 MYM-TYPE ZINC FINGER.

CC -!- CAUTION: Ref.5 sequence differs from that shown due to a

CC frameshift in position 330.

CC -!- CAUTION: Ref.6 sequence differs from that shown due to frameshifts

CC in positions 330, 966, 1009 and 1017.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL; Y13472; CAA73875.1; -

DR EMBL; AJ224901; CAA12204.1; -

DR EMBL; AJ007676; CAA07604.1; -

DR EMBL; AJ007677; CAA07604.1; JOINED.

DR EMBL; AJ007678; CAA07604.1; JOINED.

DR EMBL; AJ007679; CAA07604.1; JOINED.

DR EMBL; AJ007680; CAA07604.1; JOINED.

DR EMBL; AJ007681; CAA07604.1; JOINED.

DR EMBL; AJ007682; CAA07604.1; JOINED.

DR EMBL; AJ007683; CAA07604.1; JOINED.

DR EMBL; AJ007684; CAA07604.1; JOINED.

DR EMBL; AJ007685; CAA07604.1; JOINED.

DR EMBL; AJ007686; CAA07604.1; JOINED.

DR EMBL; AJ007687; CAA07604.1; JOINED.

DR EMBL; AJ007688; CAA07604.1; JOINED.

DR EMBL; AJ007689; CAA07604.1; JOINED.

DR EMBL; AJ007690; CAA07604.1; JOINED.

DR EMBL; AJ007691; CAA07604.1; JOINED.

DR EMBL; AJ007692; CAA07604.1; JOINED.

DR EMBL; AJ007693; CAA07604.1; JOINED.

DR EMBL; AJ007694; CAA07604.1; JOINED.

DR EMBL; AJ007695; CAA07604.1; JOINED.

DR EMBL; AJ007696; CAA07604.1; JOINED.

DR EMBL; AL137119; CAC42467.1; -

DR EMBL; AL138688; CAC16956.1; -

DR EMBL; AF060181; AAC23591.1; ALT_FRAME.

DR EMBL; AF035374; AAB88464.1; ALT_FRAME.

DR EMBL; AF012126; AAC01561.1; -

DR MIM; 602221; -

KW Transcription regulation; Nuclear protein; Chromosomal translocation;

Repeat; Zinc-finger.

FT ZN_FING 331 409 MYM-TYPE 1 (POTENTIAL).

FT ZN_FING 425 502 MYM-TYPE 2 (POTENTIAL).

FT ZN_FING 536 616 MYM-TYPE 3 (POTENTIAL).

FT ZN_FING 639 712 MYM-TYPE 4 (POTENTIAL).

FT ZN_FING 724 799 MYM-TYPE 5 (POTENTIAL).

FT SITE 913 914 BREAKPOINT FOR TRANSLATION TO FORM ZNF198-EGFR1.

FT CONFLICT 388 416 VAOYDSSSEFQECSTCLSLYEDKQNP -> GWLKWIQV

FT CONFLICT 657 659 NKV -> ASL (IN REF. 4).

FT CONFLICT 736 736 K -> G (IN REF. 1).

FT CONFLICT 766 767 MISSING (IN REF. 4).

FT CONFLICT 967 967 MISSING (IN REF. 6).

FT CONFLICT 1009 1010 DF -> IS (IN REF. 6).

FT CONFLICT 1016 1016 MISSING (IN REF. 6).

SQ SEQUENCE 1377 AA; 154910 MW; 2652D4C766492FF9 CRC64;

Query Match 2.8%; Score 110.5; DB 1; Length 1377;

Best Local Similarity 23.4%; Pred. No. 8.9;

Matches 93; Conservative 47; Mismatches 148; Indels 109; Gaps 22;

QY 342 PALVSSYAAVTSLSGNSQEAHVYRSKGAACAAFLSNYDSRYSVKVTQN----- 389

DB 16 PVLGSGTAMATSL-----TNGNSPSPGANPLVRSNK-----FONSSVEDDDVVFI 63

QY 390 RPYNLPPHSISILPCKTAVNTAQVNS-QSSSIKWTAGGSLWSQSYN-EETPTADSD 447

DB 64 EPVQPPSPVPVADQRTITTSKNEELQNDKSKITPSSKELASOKGVSSETIVIDDEE 123

QY 448 TLTANGLWEQ-----KNVTR--DSSDYLTWMTNVTNVIASNEGFKNKGDVYL- 491

DB 124 DMETNQGEKNSNFIERPPETKNTNDVDETSFSRSKVNAGMNGSGITTEPDSEIQ 183

QY 492 ----TVMSAGHVLHVYVNGKLSGTVYGTLDNPKLTYSGNVKLR-----AGINKISLSVSV 543

DB 184 IANVTTLTETG-----VSSVNDGQLEN---TDGRDMLMLITHVTSQNTNLGDVSN 230

QY 544 GL--PNVGVDHTWNAV-----LGPVTLISGLNEGRNLAKQKWSYKVLKGSLSLH 594

DB 231 GLQSNFNGVNIQTYPSTTSQTKTGVP-----FNPGRMNVAGD-----VFQNGESATHH 280

QY 595 SLSSG--SSSVEWVRSLSMAQKPLTWYKATFNAPCGNDPLALDMSMGKGQIWIINGEGVG 652

DB 281 NPDWSISOSASEPRN-----QKQP-----GVDSLS-PVASLPK-QIFQPSVQQQ 322

QY 653 RHPFGYIAQDCSKCSYAG-TFNEKK-----CQTNC 682

DB 323 PTKPVKVTCAKCKPLQKGQTAYQRKGSALHFCSTTC 359

Search completed: June 24, 2002, 20:42:23

Job time: 230 sec


GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2002, 20:36:19 ; Search time 22.94 Seconds
(without alignments)
3032.636 Million cell updates/sec

Title: US-09-701-868-11
Perfect score: 3924
Sequence: 1 MLRTNVLLLVICLLDFSS.....VVEFMGNPTGISLVRSR 724
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues
Total number of hits satisfying chosen parameters: 283138
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : 
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3924	100.0	724	2 T04340	beta-galactosidase
2	3043	77.5	731	2 T17002	probable beta-gala
3	2922.5	74.5	835	2 T06590	probable beta-gala
4	2863	73.0	832	2 S41889	beta-galactosidase
5	2838	72.3	729	2 T04269	probable beta-gala
6	2797	71.3	727	2 T47541	beta-galactosidase
7	2676	68.2	731	2 S16595	gene CARSR12 prote
8	2608	66.5	853	2 B85429	beta-galactosidase
9	2572	65.5	853	2 T04600	probable beta-gala
10	2259.5	57.6	839	2 C84685	probable beta-gala
11	2035.5	51.9	895	2 T00787	probable beta-gala
12	1909.5	48.7	828	2 S52393	beta-galactosidase
13	1647.5	42.0	780	2 D96803	probable beta-gala
14	1642	41.8	832	2 B84543	beta-galactosidase
15	1410.5	35.9	831	2 T05771	beta-galactosidase
16	1380.5	35.2	1036	2 T05687	beta-galactosidase
17	578	14.7	586	2 JC5618	beta-galactosidase
18	547	13.9	595	2 H95006	beta-galactosidase
19	543	13.8	595	2 C97879	beta-galactosidase
20	525.5	13.4	598	2 A57249	beta-galactosidase
21	518	13.2	612	2 B82756	beta-galactosidase
22	513.5	13.1	595	2 T29434	beta-galactosidase
23	495	12.6	677	2 A32611	beta-galactosidase
24	479.5	12.2	647	2 A37086	beta-galactosidase
25	429	10.9	616	2 C96755	Similar to acid be
26	413	10.5	746	2 T24978	hypothetical prote
27	393	10.0	982	2 A97210	beta galactosidase
28	365.5	10.0	1006	2 T31685	beta-galactosidase
29	365.5	9.3	681	2 T33381	hypothetical prote

30	259.5	6.9	778	2 B71164	probable beta-gala
31	265.5	6.8	787	2 C75068	probable beta-gala
32	227.5	5.8	546	2 B32688	beta-galactosidase
33	223	5.7	171	2 S37748	beta-galactosidase
34	146.5	3.7	1455	1 A48925	mannose receptor p
35	131	3.3	374	2 JC2124	major allergen Cry
36	129.5	3.3	1661	2 T31330	head-activator bin
37	128	3.3	2204	2 A70524	probable PPE prote
38	126.5	3.2	2902	2 C71953	toxin-like outer m
39	125	3.2	374	2 JC2123	major allergen Cry
40	120	3.1	1932	2 S53409	probable membrane
41	120	3.1	2154	2 P83088	hypothetical prote
42	119	3.0	686	2 AH0104	beta-galactosidase
43	118.5	3.0	934	2 E64235	hypothetical prote
44	118	3.0	1214	2 T30941	DNA polymerase - c
45	117.5	3.0	1449	2 T30552	glucosyltransferas

ALIGNMENTS

RESULT 1

T04340

beta-galactosidase (EC 3.2.1.23) II precursor - tomato

C:Species: Lycopersicon esculentum (tomato)

C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 19-May-2000

C:Accession: T04340

R:Smith, D.L.; Starrett, D.A.; Gross, K.C.

Plant Physiol. 117, 417-423, 1998

A:Title: A gene coding for tomato fruit beta-galactosidase II is expressed during fru

A:Reference number: Z15296; MUID:98289087

A:Accession: T04340

A:Status: translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-724 <SMT>

A:Cross-references: EMBL:AF020390; NID:g3298985; PIDN:AAC25984.1; PID:g3298986

A:Experimental source: strain Rutgers; tissue-type tomato fruit

C:Genetics:

A:Gene: Bgal4

C:Superfamily: beta-galactosidase bga

C:Keywords: glycosidase; hydrolase

F:1-23/Domain: signal sequence #status predicted <SIG>

F:24-724/Product: beta-galactosidase II #status predicted <MAT>

```
QY 361 HYVRSKGACAAFLSNYDSRYSVKVTQFNRPYNLPWPWSISILPDCCKTAVYNTAQNVSOGS 420
|||||
Db 361 HYVRSKGACAAFLSNYDSRYSVKVTQFNRPYNLPWPWSISILPDCCKTAVYNTAQNVSOGS 420
QY 421 STKMTYPAGGLSWOSVNEETPTADSDTLTANGLWEQKNVTRDSSDYLWYMTNVAIASNE 480
|||||
Db 421 STKMTYPAGGLSWOSVNEETPTADSDTLTANGLWEQKNVTRDSSDYLWYMTNVAIASNE 480
QY 481 GFLKNGKDPYLVMSAGHVLHVFGVNGKLSGTYGTLDPNPKLTYSGNVKLRAGINKISLIS 540
|||||
Db 481 GFLKNGKDPYLVMSAGHVLHVFGVNGKLSGTYGTLDPNPKLTYSGNVKLRAGINKISLIS 540
QY 541 VSVGLPNVGVHYDTNWAGVLGPTVTLISGLNEGRNLAKOKWSYKVLKGESLSLHLSGSS 600
|||||
Db 541 VSVGLPNVGVHYDTNWAGVLGPTVTLISGLNEGRNLAKOKWSYKVLKGESLSLHLSGSS 600
QY 601 SVEWYRGLSMAOKQPLTWYKATFNAPGNDPLALDMSMGKQIWINEGVGRHWPYIA 660
|||||
Db 601 SVEWYRGLSMAOKQPLTWYKATFNAPGNDPLALDMSMGKQIWINEGVGRHWPYIA 660
QY 661 QGDCSKSVAGTFNEKKCOTNCGOPSORWYHVPWSLKPSPGNLLVVFPEWGNPTGISLV 720
|||||
Db 661 QGDCSKSVAGTFNEKKCOTNCGOPSORWYHVPWSLKPSPGNLLVVFPEWGNPTGISLV 720
QY 721 RRSR 724
|||||
Db 721 RRSR 724

RESULT 2
TI7002
probable beta-galactosidase (EC 3.2.1.23) precursor - apple tree
C:Species: Malus domestica (apple tree)
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 21-Jul-2000
C:Accession: TI7002
R:Ross, G.S.; Wegryn, T.; Macrae, E.A.; Redgwell, R.J.
Plant Physiol. 106, 521-528, 1994
A:Title: Apple beta-galactosidase. Activity against cell wall polysaccharides and charac
A:Reference number: 218645; MUID:95083752
A:Accession: TI7002
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-731 <ROS>
A:Cross-references: EMBL:L29451; NID:9507277; PIDN:AAAG62324.1; PID:9507278
A:Experimental source: strain Granny Smith; cortical
C:Superfamily: beta-galactosidase bga
C:Keywords: glycosidase; hydrolase
F:1-23/domain: signal sequence #status predicted <SIG>
F:24-731/product: probable beta-galactosidase #status predicted <MAT>

Query Match 77.5%; Score 3043; DB 2; Length 731;
Best Local Similarity 74.9%; Pred. No. 7.3e-204;
Matches 536; Conservative 96; Mismatches 80; Indels 4; Gaps 2;

QY 7 LLLLVICLLDFFSSVKASVSDRAIIINGKRKILISGSIHYPRSTPQMPDLIQAKDG 66
|||||
Db 11 ILLLFSCI--FSAASASVSDHKAIIINGKRKILISGSIHYPRSTPQMPDLIQAKDG 67
QY 67 GLDVETVFWNGHPSGKYNFEGRYDLVRFKMWQVAGLYVNLRIIGPYCAEWNFGGF 126
|||||
Db 68 GLDVITQTVFWNGHPSGPNYFEERYDLVRFKILVQOEGFLVNLRIIGPYCAEWNFGGF 127
QY 127 PWLVKYVPGMEFRTNQPFVKVMOGFVKIYNMMSKSENLFSOGQPIIMAOIENEGYVPE 186
|||||
Db 128 PWLVKYVPGIAFRTDNEFPKAAQKFTKIVSMMAEKLFQTOGGPIILSQIENEGYVPE 187
QY 187 WEIGAPKAYTKWAAQMAVGLKTVGPWIMCKQEDAPDVIDTCNGFYCEGFRPNKPKYKPK 246
|||||
Db 188 WEIGAPKAYTKWAAQMAVGLDTPVWIMCKQEDAPDVIDTCNGFYCEGFRPNKPKYKPK 247
QY 247 MWTEVMTGWTKFGGPIPORPAEDIAFSVAREFVNQNGSFFNYMYHGGTNGRTSSGLFI 306
|||||
```

```
Db 248 MWTEVMTGWTEFGGAVPTRPAEDVAFSVAREIQSGGFLNYMYHGGTNGRTAGGPPM 307
QY 307 ATSYDYDAPLDDEYGLLNEPKYGHRLDLHKAIKLSPALVSSVAAVTSLSGNOEAHVRSK 366
|||||
Db 308 ATSYDYDAPLDDEYGLPREPKWGHRLDLHKAIKSCESALVSDVPSVTKLGSNOEAHVRSK 367
QY 367 SGACAAFLSNYDSRYSVKVTQFNRPYNLPWPWSISILPDCCKTAVYNTAQNVSOGS 426
|||||
Db 368 SD-CAAFLANYDAKYSVKVSGGGQYDLPPWPSISILPDCCKTEVNTAKVSSOSSQVQWTP 426
QY 427 AGGGLSWOSVNEETPTADSDTLTANGLWEQKNVTRDSSDYLWYMTNVAIASNEGFLKNG 486
|||||
Db 427 VHSFGPWOSFIBETTSDETDTTLLDGLYEQINIRDTTDLWYMTDITIGSDEAFKNG 486
QY 487 KDPYLTVMASAGHVLHVFGVNGKLSGTYGTLDPNPKLTYSGNVKLRAGINKISLISVSVGLP 546
|||||
Db 487 KSPLUTIFSAHALNANFINGLQSGTYGSLNPKLSFQNVNLRSGINKLALLSISVGLP 546
QY 547 NVGHYDTNWAGVLGPTVTLISGLNEGRNLAKOKWSYKVLKGESLSLHLSGSSSVEWYR 606
|||||
Db 547 NVGTHFTWNAGVLGPTITLKLNSGTWDMGSKWTKYTKLKGALGLHTVTGSSSVEWVE 606
QY 607 GSLMAOKQPLTWYKATFNAPGNDPLALDMSMGKQIWINEGVGRHWPYIAQDCSK 666
|||||
Db 607 GPSMAEKQPLTWYKATFNAPGNDPLALDMSMGKQIWINESQSVGRHWPYIARGSCGD 666
QY 667 CSYAGTFNEKKCOTNCGOPSORWYHVPWSLKPSPGNLLVVFPEWGNPTGISLVRR 722
|||||
Db 667 CSYAGTYDDKKCRTHCGEPPSQRWYHVPWSLTPGNTGNLLVVFPEWGNPTGISLVRR 722

RESULT 3
T06590
probable beta-galactosidase (EC 3.2.1.23) - tomato
C:Species: Lycopersicon esculentum (tomato)
C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 19-May-2000
C:Accession: T06590
R:Carey, A.T.; Holt, K.; Picard, S.; Wilde, R.; Tucker, G.A.; Bird, C.R.; Schuch, W.;
Plant Physiol. 108, 1099-1107, 1995
A:Title: Tomato exo-(1-4)-beta-D-galactanase: isolation, changes during ripening in n
A:Reference number: 215780; MUID:95357407
A:Accession: T06590
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-835 <CAR>
A:Cross-references: EMBL:X83854; NID:971484; PIDN:CAA58734.1; PID:971485
A:Experimental source: cultivar Ailsa Craig; pericarp
C:Superfamily: beta-galactosidase bga
C:Keywords: glycosidase; hydrolase

Query Match 74.5%; Score 2922.5; DB 2; Length 835;
Best Local Similarity 72.2%; Pred. No. 2.2e-195;
Matches 518; Conservative 94; Mismatches 102; Indels 3; Gaps 2;

QY 6 VLLLVICLLDFFSSVKASVSDRAIIINGKRKILISGSIHYPRSTPQMPDLIQAKD 65
|||||
Db 7 MLLMLLICL--WVSGIASVSDHKAIIINGKRKILISGSIHYPRSTPQMPDLIQAKD 64
QY 66 GLDVETVFWNGHPSGKYNFEGRYDLVRFKMWQVAGLYVNLRIIGPYCAEWNFGG 125
|||||
Db 65 GGDVDIQTIVFWNGHPSGKYNFEGRYDLVRFKMWQVAGLYVNLRIIGPYCAEWNFGG 124
QY 126 FPWLKYVPGMEFRTNQPFVKVMOGFVKIYNMMSKSENLFSOGQPIIMAOIENEGYV 185
|||||
Db 125 FPWLKYVPGISFRTNNEFPKAAQKFTKIVDMMAKELIETOGGPIILSQIENEGYV 184
QY 186 EWEIGAPKAYTKWAAQMAVGLKTVGPWIMCKQEDAPDVIDTCNGFYCEGFRPNKPKYKPK 245
|||||
Db 185 EWEIGAPKAYTKWAAQMAVGLDTPVWIMCKQEDAPDVIDTCNGFYCEGFRPNKPKYKPK 244
QY 246 MWTEVMTGWTKFGGPIPORPAEDIAFSVAREFVNQNGSFFNYMYHGGTNGRTSSGLFI 305
|||||
```

Db 245 KMTAEWTANWTEFGGPPVDPYPAEDMAFAVARFIQTGGSFINYMYHGGTNGRSTGGGPF 304
Qy 306 IATSYDYDAPLDEYGLLNEPKYGLHRLDKAHLKALSEPALVSSYAAVTSIGSNOEAHVRS 365
Db 305 IATSYDYDAPLDEYGLLNEPKYGLHRLDKAHLKALSEPALVSDPTVTSIGNQEARVKS 364
Qy 366 KSGACAFLSNYSRYSVKVFQNRPNLPPWSISILPDKCTAVYNTAQVNSOSSIKMT 425
Db 365 ESGACAFLSNYSRYSVKVFQNRPNLPPWSISILPDKCTAVYNTARVQAQSMKMT 424
Qy 426 PAGGGLSWQSYNEETPTADSDTLTANGLWEQKNVTRDSSDYLWYMTNVTNVIASNEGFLN 485
Db 425 PVSRGFSHESFEDDAESHED-DFTFVGLLEQINTRDVSDFLWMTDIEDPTGEFLNS 483
Qy 486 GKDPYLTVMASGHLVHVFNKLSGTVYGLDNPDKLTSGNPKLTKRAGINKISLLSVSGL 545
Db 484 GNPWLTVMASGHLVHVFNKLSGTVYGLDNPDKLTSGNPKLTKRAGINKISLLSVSGL 543
Qy 546 PNVGHHYDTWAGVLPVTLTSLNGESRNLAQKWSYKVLKGESLSLHSLSGSSVEMV 605
Db 544 PNVGHHYDTWAGVLPVTLTSLNGESRNLAQKWSYKVLKGESLSLHSLSGSSVEMV 603
Qy 606 RGSILMAQKQPLTWYKATFNAPGNDPLALDMSMGCKGOIWIINGEGVGRHWPYIAQGDPS 665
Db 604 EGSILVAQKQPLTWYKATFNAPGNDPLALDMSMGCKGOIWIINGEGVGRHWPYIAQGDPS 663
Qy 666 KCSYAGTFNEKKCOTNCGQPSQRYHVPWSLKPNSGILLVFEWGGNPTGISLVRR 722
Db 664 VCNVTGWDFEKKCLTNCGEGSQRYHVPWSLKPNSGILLVFEWGGNPTGISLVRR 720

RESULT 4
S41889
beta-galactosidase (EC 3.2.1.23) - garden asparagus
C:Species: Asparagus officinalis (garden asparagus)
C:Date: 20-May-1994 #sequence_revision 10-Nov-1995 #text_change 19-May-2000
C:Accession: S41889
R:King, G.A.; Davies, K.M.; Stewart, R.G.; Borst, W.M.
submitted to the EMBL Data Library, January 1994
A:Description: Identification and characterization of cDNA clones for asparagus mRNAs
A:Reference number: S41889
A:Accession: S41889
A:Molecule type: mRNA
A:Residues: 1-832 <KIN>
A:Cross-references: EMBL:X77319; NID:g452711; PIDN:CAA54525.1; PID:g452712
A:Superfamily: beta-galactosidase bga
C:Keywords: glycosidase; hydrolase

Query Match 73.0%; Score 2863; DB 2; Length 832;
Best Local Similarity 71.4%; Pred. No. 3e-191;
Matches 514; Conservative 96; Mismatches 102; Indels 8; Gaps 5;

Qy 6 VLLLVICLLDFFS--SVKASVYDDRAIIINGKRKILISGSIHYPRSTPOMWPDLOKA 63
Db 6 VLLLVICLLDFFS--SVKASVYDDRAIIINGKRKILISGSIHYPRSTPOMWPDLOKA 65
Qy 64 KDGGLDVLTETVYFVWNGHEPSPCKYFEGRYDLVRFKVMQVAGLVNLRIGPYCAEWNF 123
Db 66 KDGGLDVLTETVYFVWNGHEPSPCKYFEGRYDLVRFKVMQVAGLVNLRIGPYCAEWNF 125
Qy 124 GGFPPWLKYVPGMEFRTNPNQPFKVMQGVOKIYNNMKSLENLFESQGGPIIQAQENY 183
Db 126 GGFPPWLKYVPGMEFRTNPNQPFKVMQGVOKIYNNMKSLENLFESQGGPIIQAQENY 185
Qy 184 PVEWIGAPGKAYTKWAQMAVGLKTVPMWCKQEDAPDPVIDTCNCFYCEGFRPNKPY 243
Db 186 PVEWIGAPGKAYTKWAQMAVGLKTVPMWCKQEDAPDPVIDTCNCFYCEGFRPNKPY 245
Qy 244 KPMWTEWGTWYTKFGGPIQORPAEDIAFSVAREVQNNGSFFNYMYHGGTNGRSTSSG 303
Db 246 KPMWTEWGTWYTKFGGPIQORPAEDIAFSVAREVQNNGSFFNYMYHGGTNGRSTSSG 305

Qy 304 LFIATSYDYDAPLDEYGLLNEPKYGLHRLDKAHLKALSEPALVSSYAAVTSIGSNOEAHVY 363
Db 306 LFIATSYDYDAPLDEYGLLNEPKYGLHRLDKAHLKALSEPALVSGEPTITSILQONQESVY 365
Qy 364 RSKSACAAFLSNYSRYSVKVFQNRPNLPPWSISILPDKCTAVYNTAQVNSOSSSIK 423
Db 366 RSKSACAAFLSNYSRYSVKVFQNRPNLPPWSISILPDKCTAVYNTARVQAQTTMK 424
Qy 424 MTPAGGGLSWQSYNEETPTADSDTLTANGLWEQKNVTRDSSDYLWYMTNVTNVIASNEGFL 483
Db 425 MTPAGGGLSWQSYNEETPTADSDTLTANGLWEQKNVTRDSSDYLWYMTNVTNVIASNEGFL 482
Qy 484 KNGKDPYLTVMASGHLVHVFNKLSGTVYGLDNPDKLTSGNPKLTKRAGINKISLLSVS 543
Db 483 KNGKDPYLTVMASGHLVHVFNKLSGTVYGLDNPDKLTSGNPKLTKRAGINKISLLSVS 542
Qy 544 GLPNVGHYDTWAGVLPVTLTSLNGESRNLAQKWSYKVLKGESLSLHSLSGSSSVE 603
Db 543 GLPNVGHYDTWAGVLPVTLTSLNGESRNLAQKWSYKVLKGESLSLHSLSGSSSVE 602
Qy 604 WYRGSILMAQKQPLTWYKATFNAPGNDPLALDMSMGCKGOIWIINGEGVGRHWPYIAQGD 663
Db 603 WYRGSILMAQKQPLTWYKATFNAPGNDPLALDMSMGCKGOIWIINGEGVGRHWPYIAQGD 659
Qy 664 CSKCSYAGTFNEKKCOTNCGQPSQRYHVPWSLKPNSGILLVFEWGGNPTGISLVRR 723
Db 660 CSKCSYAGTFNEKKCOTNCGQPSQRYHVPWSLKPNSGILLVFEWGGNPTGISLVRR 719

RESULT 5
T04269
probable beta-galactosidase (EC 3.2.1.23) - Arabidopsis thaliana
N:Alternate names: protein F20B18.250
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 30-Apr-1999 #sequence_revision 30-Apr-1999 #text_change 19-May-2000
C:Accession: T04269
R:Bevan, M.; Rose, M.; Hempel, S.; Entian, K.D.; Hoheisel, J.; Mewes, H.W.; Mayer, K.
submitted to the Protein Sequence Database, March 1999
A:Reference number: 215263
A:Accession: T04269
A:Molecule type: DNA
A:Residues: 1-729 <BEV>
A:Cross-references: EMBL:AL049483
A:Experimental source: cultivar Columbia; BAC clone F20B18
C:Genetics:
A:Map position: 4
A:Introns: 58/3; 90/3; 128/2; 150/3; 181/3; 229/3; 259/2; 294/3; 323/1; 362/3; 416/3;
A:Note: F20B18.250
A:Superfamily: beta-galactosidase bga
C:Keywords: glycosidase; hydrolase

Query Match 72.3%; Score 2838; DB 2; Length 729;
Best Local Similarity 69.7%; Pred. No. 1.4e-189;
Matches 502; Conservative 104; Mismatches 108; Indels 6; Gaps 5;

Qy 7 LLLLVICLLDFFS--SVKASVYDDRAIIINGKRKILISGSIHYPRSTPOMWPDLOKADG 66
Db 11 LLLLVICLLDFFS--SVKASVYDDRAIIINGKRKILISGSIHYPRSTPOMWPDLOKADG 70
Qy 67 GLDVIETVYFVWNGHEPSPCKYFEGRYDLVRFKVMQVAGLVNLRIGPYCAEWNFGF 126
Db 71 GLDVIETVYFVWNGHEPSPCKYFEGRYDLVRFKVMQVAGLVNLRIGPYCAEWNFGF 130
Qy 127 PWLKYVPGMEFRTNPNQPFKVMQGVOKIYNNMKSLENLFESQGGPIIQAQENYGPVE 186
Db 131 PWLKYVPGMEFRTNPNQPFKVMQGVOKIYNNMKSLENLFESQGGPIIQAQENYGPVE 190
Qy 187 WEIGAPGKAYTKWAQMAVGLKTVPMWCKQEDAPDPVIDTCNCFYCEGFRPNKPYKPK 246
Db 191 WEIGAPGKAYTKWAQMAVGLKTVPMWCKQEDAPDPVIDTCNCFYCEGFRPNKPYKPK 250
Qy 247 MWTEWGTWYTKFGGPIQORPAEDIAFSVAREVQNNGSFFNYMYHGGTNGRSTSSGLFI 306

Db	192	EWEI	GA	PKAYTHWA	AQAQASINAGVPIWIMCKODSDVDPNVIDTCNGFVCPKDKSK	251
Qy	245	PKMT	VE	TGWYTKFGGPI	QPAEDIAFVSARFVQNNGSFFNYMYHGGTNFGRSSGL	304
Db	252	PKMT	EN	TGWYTEXGKVPY	PAEDVAFVSARFIQNGSGFMNYMFHGCTNF-EITAGR	310
Qy	305	FIAT	SY	DYDAPLDEYGLL	NEPKYGHRLDHKAIKLSEPALVSSYAATVSLGSGNQEAHVYR	364
Db	311	FVST	SY	DYDAPLDEYGL	PREPKYTHLKNJHKAIKMCEPALVSSDAKVTNLGSGNQEAHVYS	370
Qy	365	SKSG	CA	AAFLSNYS	SRYSVKVTFQNRPNYLPWPWSISILPDCKTAVYNTAQVNSQSSSI--	422
Db	371	SNSG	CA	AAFLANYP	KWSVKVTFSGMEPELPAWSISILPDCKEVYNTARVNEPSPKLHS	430
Qy	423	KMT	P	AGGLS	WQSYNEETPTADDSDTLTANGLWEQKNVTRDSSDYLWYMTNVNIASNEGF	482
Db	431	KMT	P	ISNLN	WQSYSEVDEPTADSPGTFREKLEQINMTWDSYLLWYMTDVVLDCNEG	490
Qy	483	LKNG	K	DPLYTVMS	AGHVLHVFVNGKLSGTVYGTLDNPKLTYSGNVKLRAGINKISLLSVS	542
Db	491	LKKG	D	EWLTVNS	AGHVLHVFVNGQLQGHAYGSLAKPOLTFQKVKMTAGVNRISLLSAV	550
Qy	543	VGLP	N	VGHYDTN	AGVLGPVTLTSLGNEGSRNLAKQKSYKVLCKGESLHLSLGSQSSSV	602
Db	551	VGLA	N	VGHWFERY	NOGVLPVTLTSLGNEGTRDLTWQYWSYKIGTKGEEQQVYVNSGGSSHV	610
Qy	603	EWVR	S	LMAQKQPL	TYKATFNAPGNDPLALDMSMGKGQIWIINGEGVGRHWPGYIAQG	662
Db	611	QW--	--	GPPAW	KQPLVWYKTTFDAPGNDPLALDLSMGKGQAWINQSIGRHWNNIAKG	667
Qy	663	DCS-	KCS	YAGTFEN	EKKQTCNCCQPSQRWYHVPBSWLKPSGNLLVYFEEWGGNPTGISLVR	721
Db	668	SCND	NC	NYAGTYTETK	CLSDCKSGKQKWKYHVPBSWLQPRGNLLVYFEEWGGDTKKVSLVK	727
Qy	722	RS	723			
Db	728	RT	729			
RESULT	8					
B85429						
beta-galactosidase like protein [imported] - Arabidopsis thaliana						
C:Species: Arabidopsis thaliana (mouse-ear cress)						
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 02-Mar-2001						
C:Accession: B85429						
R:anonymous, The European Union Arabidopsis Genome Sequencing Consortium, The Nature 402, 769-777, 1999						
A:Title: Sequence and analysis of chromosome 4 of the plant Arabidopsis thaliana						
A:Reference number: B85001; MUID:20083488						
A:Accession: B85429						
A:Status: preliminary						
A:Molecule type: DNA						
A:Residues: 1-853 <STO>						
A:Cross-references: GB:NC_001268; NID:g7270584; PIDN:CAB80302.1; GSPDB:GN00140						
C:Genetics:						
A:Gene: At4g36360						
A:Map position: 4						
C:Superfamily: beta-galactosidase bqa						

[illegible]

Query Match	66.5%	Score	2608;	DB	2;	Length	853;
Best Local Similarity	63.1%	Pred. NO.	1.8e-173;				
Matches	457;	Conservative	119;	Mismatches	136;	Indels	12;
						Gaps	5;

QY 62 KAKDGLDVIETVFNWGHSPGKYNFEGRYDLVRFIKMVORAGLYVNLRTGIPYVCAEW 121
Db 70 KAKDGDIDVIETVFNWHLHSPGKYDFEGRNDLRFVFKTIHKAGLYAHLRIGPYVCAEW 129
QY 122 NFGGPPVWLKYPGMEFRNNOFPKVMQGFQKIVNMKSENLFESQGGPIIQAQIENE 181
Db 130 NFGGPPVWLKYPGSIFFRDNPEFPKRMKGFTERIVELMKSENLFESQGGPIIISOIENE 189
QY 182 YGPVEWIEGAPKAYTKWAAQMAVGLKTGVPWIMCKOEADPDVDTGNGFYCEGFRPNK 241
Db 190 YGRQGLIGAEGHNTWTAAKNAIATETGVPWIMCKEDDAPDVINTGNGFYCDSEFAPNK 249
QY 242 PYKPKMTWETVWTGWTYKFGGPIQORPAEDIAFSVARFVQNNGSFFNYMYHGGTNGFRGTS 301
Db 250 PYKPLIWTAEWSGWTETFGGPHHRPQDLAFGVARFQKGGSFVNYMYHGGTNGFRGTA 309
QY 302 SGLFATSYDYDAPLDEYGLLNEPKYGHLDLHKAIKLSEPALVSSYAAVTSLSGNOE-- 359
Db 310 GGFVVTTSYDYDAPIDEYGLIRQPKYGHLDLHKAIKLSEPALVSSYAAVTSLSGNOE-- 369
QY 360 -----AHVYRSKGACAAFLSNYSRYSVKVTQFQNRPNLPPWISILPDCCKTAYNTA 413
Db 370 IYERFAHYVSAESGCSAFLANYDTESARVLFNNVHNLPPWISILPDCRNVAFNTA 429
QY 414 QVNSOSSIKMTPAGGGLSWQSYNEETPTADDSDTLTANGLWEQKNVTRDSSDYLWYMTN 473
Db 430 KVSN-----FQWESYLEDLSSLDSSSTFTTHGLLEQINVTROTSDYLWYMTS 476
QY 474 VNIASNEGLKNDKDPYLTVMSAGHVLHVFNGLKSGIVYGTGLDNPDKLTYSNVKLKRAI 533
Db 477 VDIGSESEFLHGGELPTLIQSTGHAVHIFVNGQLSGAFGTRNRRFTYQCKINLHSGT 536
QY 534 NKISLLSVSVGLPNVGVHYDTWNAAGVLPVTLNGLNEGRNLAKQWSYKVLKGSLSL 593
Db 537 NRIALLSVAVGLPNVGGHESNWTGILGPVALHLSQGMKDLWSQWRYQVGLKGEAMNL 596
QY 594 HSLSGSSVWVYRGLSMAQK-OPLTWYKATFNAPGNNDPLALDMSMGKQIWIINGEGV 652
Db 597 APTNTPSIGWMDASLTVOKPQPLTWHTYFDPAEGNEPLALDMEGCKGQIWIINGESIG 656
QY 653 RHWPGYIAQDCSKSYAGTFNEKKCOTNCGQPSORWYHVPWSLKPNSNLLVFEWGG 712
Db 657 RYWTAF-ATGDCSHCSYGTYYKPNKQTCGCGQPTQRYWYHVPRAWLKPSONLLVIFEELGG 715
QY 713 NPTGISLVRRS 723
Db 716 NPSTVSLKRS 726
RESULT 10
C84685
probable beta-galactosidase [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 16-Feb-2001
C:Accession: C84685
R.; Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon,
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter,
Nature 402, 761-769, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: A84420; MUID:20083487
A:Accession: C84685
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-839 <STO>
A:Cross-references: GB:AE002093; NID:q4510395; PIDN:AAD21482.1; GSPDB:GN00139
C:Genetics:
A:Gene: At2g28470
A:Map position: 2
C:Superfamily: beta-galactosidase bga

Query Match 57.6%; Score 2259.5; DB 2; Length 839;
Best Local Similarity 57.1%; Pred. No. 3.2e-149;
Matches 418; Conservative 111; Mismatches 176; Indels 27; Gaps 10;
QY 3 RTNLLLLVICLLDFESSVKASVSYYDDRAIINGRKILISGSIHYPRSTPOMWPDLOK 62
Db 6 KHEMILLILIV--VAATAANVTYDHRALVDGRRKVLISGSIHYPRSTPEMPELIQ 63
QY 63 AKDGLDVIETVFNWGHSPGKYNFEGRYDLVRFIKMVORAGLYVNLRTGIPYVCAEW 122
Db 64 SKDGLDVIETVFNWGHSPGKYNFEGRYDLVRFIKMVORAGLYVNLRTGIPYVCAEW 123
QY 123 FGGFPVWLKYPGMEFRNNOFPKVMQGFQKIVNMKSENLFESQGGPIIQAQIENEY 182
Db 124 YGGFPVWLHFPVGIKFRDNEPFEKMQRTTKIVDLMKQEKLYASQGGPIIISOIENEY 183
QY 193 GPEWIEGAPKAYTKWAAQMAVGLKTGVPWIMCKOEADPDVDTGNGFYCEGFRPNK 242
Db 184 GNIDSAYGAAKSYIKWSASMAISLDTGVPWIMCQOTDAPDPMINTGNGFYCDQFTPSN 243
QY 243 YPKMWTETVWTGWTYKFGGPIQORPAEDIAFSVARFVQNNGSFFNYMYHGGTNGFRGTS 302
Db 244 NKPKMTENWSGFWLFGGDPSPYRPVEDLAFVAFYQGGTFQNYMYHGGTNGFRGTS 303
QY 303 GLFIATSYDYDAPLDEYGLLNEPKYGHLDLHKAIKLSEPALVSSYAAVTSLSGNOEAVH 362
Db 304 GPLISTSYDYDAPIDEYGLLRQPKWHLRDLHKAIKLCEADALIAITDPTTISLGSLEAAV 363
QY 363 YRSKSGACAAFLSNYSRYSVKVTQFQNRPNLPPWISILPDCCKTAVNTAQVNSQSSI 422
Db 364 YTESGSCAAFLANVDTKSDATVTFNGKSYNLPANWSILPDCKNVAFTAKV--KFNSI 421
QY 423 KMTAGG-----GLSQWYNEETPTADDSDTLTANGLWEQKNVTRDSSDYLWYMTNNTA 477
Db 422 SKTPDGGSSAELGSSW-SYIKEPISKADAPLKLGLLEQINTTADKSDYLWYSLRTDIK 480
QY 478 SNEGFLKNGKDPYLTVMSAGHVLHVFNGLKSGIVYGTGLDNPDKLTYSNVKLKRAI 533
Db 481 GDETFDEGSKAVLHIESIGQVYAFINGKLKAGSGHGKQKISLDIP-----INLVGT 533
QY 534 NKISLLSVSVGLPNVGVHYDTWNAAGVLPVTLNGLNEGRNLAKQWSYKVLKGSLSL 592
Db 534 NPTIDLSVTVGLANGVAFDLVAGITGPTVLKSAKGGSSIDLASQQWYQVGLKGEDTG 593
QY 593 LHLSLGSSSVWVYRGLSMAQK-OPLTWYKATFNAPGNNDPLALDMSMGKQIWIINGEGV 652
Db 594 LATVDSS---EWVSKSPLTKQPLWYKTFDAPSGSEPAIDFTGCKGIATWVNGQSIG 650
QY 653 RHWPGYIA-QGDCSK-CSYAGTFNEKKCOTNCGQPSORWYHVPWSLKPNSNLLVFEW 710
Db 651 RYWPTSIAGNGGCTESCDYRGSYRANKCLKNGKPSQTLYHVPWSLKPNSNLLVFEW 710
QY 711 GSNPTGISLVRR 722
Db 711 GGDPTQISFATK 722
RESULT 11
T00787
probable beta-galactosidase (EC 3.2.1.23) F24L7.5 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 16-Feb-2001
C:Accession: T00787; G84737
R.; Rounsley, S.D.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes, S.M.; K
submitted to the EMBL Data Library, February 1998
A:Description: Arabidopsis thaliana chromosome II BAC F24L7 genomic sequence.
A:Reference number: Z14204
A:Accession: T00787
A:Status: translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-895 <ROU>
A:Cross-references: EMBL:AC003974; NID:g2914688; PID:g2914710
A:Experimental source: cultivar Columbia

R: Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
 M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.;
 euss, D.; Niernan, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.
 Nature 402, 761-768, 1999
 A: Title: Sequence and analysis of chromosome 2 of the plant *Arabidopsis thaliana*.
 A: Reference number: A84420; MUID:20083487
 A: Accession: G84737
 A: Status: preliminary
 A: Molecule type: DNA
 A: Residues: 1-895 <STO>
 A: Cross-references: GB:AE002093; NID:g2914710; PIDN:AAC04500.1; GSPDB:GN00139
 C: Genetics:
 A: Gene: F24L7.5; At2g32810
 A: Map position: 2
 A: Introns: 67/3; 99/3; 137/2; 159/3; 190/3; 238/3; 268/2; 303/3; 333/1; 381/3; 440/3; 51
 C: Superfamily: beta-galactosidase bga
 C: Keywords: glycosidase; hydrolase

QY 702 NLLVFEWGGNPT 715
| : : ||| ||: |
Db 713 NTLTFEEMGGDPS 726

RESULT 13

D96803

probable beta-galactosidase [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cross)

C>Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 23-Mar-2001

C:Accession: D96803

R:Theologos, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Huijar, L.
Nature 408, 816-820, 2000

A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Maithi, R.; Marziali,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A>Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A:Reference number: A86141; MUID:21016719

A:Accession: D96803

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-780 <STO>

A:Cross-references: GB:AE005173; NID:g1079481; PIDN:AAG29193.1; GSPDB:GN00141

C:Genetics:

A:Gene: F2P24.12

A:Map position: 1

C:Superfamily: beta-galactosidase bga

Query Match 42.0%; Score 1647.5; DB 2; Length 780;
Best Local Similarity 47.0%; Pred. No. 1.2e-106;
Matches 332; Conservative 102; Mismatches 205; Indels 67; Gaps 14;

QY 23 ASVSYDDRAIIINGKRKILISGSIHYPRSTPQWPDLOKAKDGLDVIETVYFVWNGHEP 82

Db 10 ANVTYDGRSLIDGHEKILFSGSIHYTRSTPQWMPSLIAKAKSGGIDVVDVTVFVWNVHEP 69

QY 83 SPKVFEGRYDLVRFIKMVORAGLYVNLRIQYICAEKNFGFPVWLKYVPGMEFRINN 142

Db 70 QOQDFDSGRDIDVRFIKVKNHGLVCLURIGFFIQGSEYSGGLFPLHNVGIVFRTDN 129

QY 143 QPFKVMQGVOKIVNMKSENLFESQGGPIIMAOIENEGYGVPEWIEGAPKAYTKWAAQ 202

Db 130 EPFKYHMKRYAKMIVKMKSENLYASQGGPIILSIOENEGYGMVGRAFRQEGKSYVKWTAK 189

QY 203 MAVGLKTGVPWIMCKQEDAPDPVIDTCNGFYC-EGFR-PNKPYKPMWTEVMTGWTKFG 260

Db 190 LAVELDTGVPWVMCKQDDAPDPLVNACNRCQGETFKGPNSPKPAIWTENWT----- 242

QY 261 GPIQORPAEDIAFVSARFVQNNGSFNNYMYHGGTFNFGTSSGLFTATSYDYDAPLDEVG 320

Db 243 ----SLSAEDIAFHAFALFTAKNGSFNNYMYHGGTFNFGNAS-QFVITSYDQAPLDEVG 297

QY 321 LLNEPKYGLRDLHKAIKLSEPALVSSAAVTSLSGNOQFAHYRSKSGACAAFLSNYDSR 380

Db 298 LLRQPKWGLKHELHAAVKLCEPPLSLGLQTTISLGLQTAFVFGKCANLCAAILVNDQKC 357

QY 381 YSVKVTQFNRPNLPWISILPDKCTAVYNTAQVNSOSSSTKMTFAGGGLS-----WQSY 436

Db 358 EST-VQFRNSSYRLSPKSVSLPDKCNFAFNATKAVNAQYNT-RTRKARONLSSPQWBEF 415

QY 437 NEETPTADDSDTLTANGLWEQKNVTRDSDYLWYTNVNIASNEGFLKNGRDPYLTVMSA 496

Db 416 TETVPSFSET-SIRGESLLEHNTTQDTSYDLWQTTFRQ-----QSEGAPSVLKVNHL 467

QY 497 GHVLHFVNGKLSGVYGTGVDLPKLTYSNGVKNLKRAGINKILLSVSGVLPNVGVHYDTWN 556

Db 468 GHALHAFVNGRFIGSMHGTFKAHRELLEKNMNLNGTNNALLSVYVGLPNSGAHLERRV 527

QY 557 AGVLGPVTLISGLNEGSRNLAOKKSYKVLKGESLSLHLSGSSSSVW--VRGSLMAQKQ 614
| : : ||| ||: |
Db 528 VGSRSVKIWNRYQLYFN--NYSWGQVGLGKEKPHVYTEDGSAKVQWQYRDS---KSQ 582
| : : ||| ||: |
QY 615 PLTWYKATENAPCGNDPLALDMSMKGQIWIWNGEGVGRHWPGYTAOGDCSKSYAGTEN 674
| : : ||| ||: |
Db 583 PLTWYKASFDTEGEDPVALNLGSMGKGAWYNGQSIAFM----- 622
| : : ||| ||: |
QY 675 EKKCOTNCGQPSQORWYHVRPSWLKPSGNLLVVF-EWEGNPTGISL 719
| : : ||| ||: |
Db 623 -----SYFRYHIPRSLKPNLNLLVLEBEREGNPLGITI 657
| : : ||| ||: |

RESULT 14

E84543

probable beta-galactosidase [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cross)

C>Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 16-Feb-2001

C:Accession: E84543

R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon,
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter
Nature 402, 761-768, 1999

A>Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.

A:Reference number: A84420; MUID:20083487

A:Accession: E84543

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-832 <STO>

A:Cross-references: GB:AE002093; NID:g4581116; PIDN:AAD24606.1; GSPDB:GN00139

C:Genetics:

A:Gene: At2g16730

A:Map position: 2

C:Superfamily: beta-galactosidase bga

Query Match 41.8%; Score 1642; DB 2; Length 832;
Best Local Similarity 45.0%; Pred. No. 3.2e-106;
Matches 321; Conservative 127; Mismatches 224; Indels 42; Gaps 13;

QY 6 VLLLVICLLDFFSSVKASVSYYDDRAIIINGKRKILISGSIHYPRSTPQWPDLOKAKD 65

Db 11 LLLVALVILSFGAL--SITYDGTSLIINGNRELLYSGSIHYPRSTPEMWNPIIKRAKQ 68

QY 66 GGLDVIETVYFVWNGHEPSPGKYNFEGRYDLVRFIKMVORAGLYVNLRIQYICAEKNFG 125

Db 69 GGLNTIQTYVFWNVHHEPEQGFNFSSRADLVKFKLIEKNGLYVTLRLGPFQAEWTHGG 128

QY 126 FPVWLKYVPGMEFRTNNOFPKVMQGVOKIVNMKSENLFESQGGPIIMAOIENEGYGPV 185

Db 129 LPYLWREVPGIFFRDNEPFEHTERYVYKVLDMMKKEKLFASQGGPIILGOIENEYSAV 188

QY 186 EWEICAPKAYTKWAAQMAVGLKTGVPWIMCKQEDAPDPVIDTCNGFYC-EGF-RPNKPY 243

Db 189 QRAYEDELNLYIKWAKSLVHMDLGIPIWMCKQNDAPDPIMNACNRCGCGDTFPGPNKDN 248

QY 244 KPMWTEVMTGWTKFGGPIQORPAEDIAFVSARFVQNNGSFNNYMYHGGTFNFGTSSG 303

Db 249 KPSLWENWTTQFRVFGDPPAQRSVEDIAYSARFESKNGTHVNYMYHGGTFNFGTSAH 308

QY 304 LFIATSYDYDAPLDEYGLLNEPKYGLHLDLHKAIKLSEPALVSSAAVTSLSGNOQEAHYV 363

Db 309 -YVTTYDDAPLDEYGLLNEPKYGLHLDLHKAIKLSEPALVSSAAVTSLSGNOQEAHYV 367

QY 364 RSK-SGACAAFLSNYDSRYSVKVTQFNRPYNLPPWISILPDKCTAVYNTAQVNSOSS- 421

Db 368 EOPGKVCACAAFLANNNTAEKIKFRGEYLIPHRISILPDKCTAVYNTGEIISHHTSR 427

QY 422 --IKMTPAGGGLSWSYNEETPTADDSDTLTANGLWEQKNVTRDSDYLWYTNVNIASN 479

Db 428 NFMKSKKANKNFDKVFVTFESVPSKIGDSFIPVELY---GLTKDESQYGWYTSFKIDN 484

QY 480 EGFLKNGKDPYLTVMSAGHVLHVFVNGKLSGTVYGTLDNPKLTYSGNVNKLRAAGINKISLL 539
Db 485 DLSKKGGKGNLRIASLGHALHVLWNGEYLNGHSGHEEKSFVQKDPVTLKEGENHLTML 544
QY 540 SVSGLPNVGVHYDTWAGVLGP--VTLSGLNGESRLAQ-KWSYKVLKLGESLSLHSL 596
Db 545 GVLTFGPDSSGYMB---HRTGPRKSVLSILGLSGTLDLTENKGNKVGMEGERLGHAE 601
QY 597 SGSSSVWVRGSLMAQKOP-LTWYKATFNAPGGNDPLALDMASMGKQIWIINGEGVGRHW 655
Db 602 EGLKKYKWERAS---GREPCMTVQTYFDAPESOSAAIRMGNGKGLIWNNGEGVGRYW 658
QY 656 PGYIAQDCSKSVAGTFNEKKQTCNGQPSQRWYHVPKSLPNSNLLVVFEE 709
Db 659 MSFLSP-----LGQPTQIEYHIPRSLPKKNLLVIFEE 692

RESULT 15
T05771
beta-galactosidase homolog M4E13.70 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 09-Apr-1999 #sequence_revision 09-Apr-1999 #text_change 02-Sep-2000
C:Accession: T05771
R:Bevan, M.; Purnelle, B.; Boutry, M.; Coffeau, A.; Hoheisel, J.; Mewes, H.W.; Mayer, K.
submitted to the Protein Sequence Database, April 1998
A:Reference number: Z15451
A:Accession: T05771
A:Molecule type: DNA
A:Residues: 1-831 <BEV>
A:Cross-references: EMBL:AL02023
A:Experimental source: cultivar Columbia; BAC clone M4E13
C:Genetics:
A:Map position: 4
A:Introns: 26/1; 83/3; 115/3; 153/2; 169/3; 217/3; 249/2; 284/3; 352/3; 408/3; 467/2; 500/3
A:Note: M4E13.70
C:Superfamily: beta-galactosidase bga

Query Match 35.9%; Score 1410.5; DB 2; Length 831;
Best Local Similarity 39.8%; Pred. No. 4.2e-90;
Matches 294; Conservative 117; Mismatches 232; Indels 95; Gaps 16;

QY 2 LRNVLLLVICLLDFSSVKA-----SVSYDDRAIIINGKRKILSGSIHYP 49
Db 19 LRKPENLTVLVLLSSSSFAAKDKKKKNKEVYDGTSLIIDGKRELLYSGSIHYP 78
QY 50 RSTPQWMPDLIQAKDGLDVIETYPWNGHEPSPGYNEGVDIVRFIKWQVRLYV 109
Db 79 RSTPEMWPSTIKRAQOGLTIQTYFVWVHEPQQGKFNFSGRADLVKFKLIKQNGMY 138
QY 110 NLRIGPVVCAEWNFSGFPVWLKYPVPGMEFTNNQPPKVMQGFVKIVNMKSENLFESQ 169
Db 139 TLRGPFIOAEWHG-----YITRDHN-----TAGAYRK----- 169
QY 170 GGPPIAQAIEIXGPVWEIGAPGKAYTKWAAQMAVGLKTGPWIMCKQEDADPDVIDTC 229
Db 170 -----IENXSAVQRAYKQDGLNYKWNALVDSMKLGPWVCKQNDAPDPMINAC 221
QY 230 NGFYC-BGF-RPNKYPKPKWTEVWGTGWYTKFGGPIQORPAEDIAFSVARFVNNGSFFN 287
Db 222 NGRHCGDTFFGPNRENKPSLWNTENWTTQFRVFGDPPTQRSVEDIAYSVARFFSKNGTHVN 281
QY 288 YMYVHGNTNFGRTSSGLFIATSYDAPLDYGLLNEPKYCHLRLHKAIKLSEPALVSS 347
Db 282 YMYTHGNTNFGRTSAH-YVTRYDYDAPLDYGLLEKPKYCHLRLHKAHLNALMLCKPLWNG 340
QY 348 YAAVTSLSGNOEAHVYRSK-SGACAAFLSNYDSRYSVKVTQFNRPYNLPPWSISILPDC 406
Db 341 QPKTEKPGKDEIRYEQPGTKCAAFANNTEAAETIKFGREYVIAPRISILPDC 400
QY 407 TAVYNTAQVNSQSSS----IKMTAGGGLSWQSYNEETPTADDSLTTLFANGLEOKNVTRD 463
Db 401 TVVYNTAQIVSQHTRNFMSKRAKFKDFKFTETLPSKLEGNYSIPVELY---GLTKD 457

QY 464 SSDYLWMTNVTNVIASNEGFLKNGKDPYLTVMSAGHVLHVFVNGKLSGTVYGTLDNPKLTY 523
Db 458 KTDYGWTTTSFKVHKHNLPTKKGVKTFVRIASLGHALHVLWNGEYLNGHSGHEEKSPVF 517
QY 524 SGNVKLRAGINKISLLSVSVGLPNVGVHYDTWAGVLGPVTLSGLNGESRLNAK-QKWSY 582
Db 518 QKVTLKAGENHLVMLGLTGFDPDSSGYMEHRYTGPQG-LSILGTSGLDLETSSKWN 576
QY 583 KVGLKGESLSLHSLSGSSSVWVRGSLMAQKOP-LTWY-----KATFNAPGGNDP 631
Db 577 KIGMEGKLGILHTEBGLKKVEWKK---FTGKAPGLTWIQKFSKCEETLQTYFDAPESVSA 633
QY 632 LALDMASMGKQIWIINGEGVGRHWPYIAQDCSKCSYAGTFNEKKQTCNGQPSQRWYH 691
Db 634 ATIRMHGMGKLIWNGEGVGRYMQSFLSP-----LGQPTQIEYH 673
QY 692 VPSRWLKPNSGNLLVVFEE 709
Db 674 IPRSLKPKKNLLVIFEE 691

Search completed: June 24, 2002, 20:38:50
Job time: 151 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2002, 20:33:54 ; Search time 14.86 Seconds
(without alignments)
1190.049 Million cell updates/sec

Title: US-09-701-868-11
Perfect score: 3924
Sequence: 1 MLRTNVLILLVLCILDFESS.....VVFEEGNGNPTGISLVRSR 724

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 231628 segs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep.*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep.*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep.*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep.*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2920	74.4	838	2	US-08-696-944-19
2	2896	73.8	730	2	US-08-696-944-2
3	2676	68.2	731	2	US-08-696-944-20
4	485	12.4	668	4	US-09-436-605-2
5	396.5	10.1	911	1	US-08-596-985-2
6	395	10.1	1007	2	US-08-551-459-4
7	146.5	3.7	1455	3	US-08-840-062-5
8	141	3.6	30	2	US-08-687-372-1
9	137	3.5	30	2	US-08-687-372-7
10	125	3.2	374	3	US-08-467-023-2
11	118	3.0	993	1	US-07-977-451-4
12	118	3.0	993	1	US-08-252-517-4
13	118	3.0	993	1	US-07-906-397A-4
14	118	3.0	993	1	US-08-222-299-4
15	118	3.0	993	1	US-08-601-891-4
16	118	3.0	993	2	US-09-021-324-4
17	118	3.0	993	2	US-08-434-878-4
18	118	3.0	993	5	PCT-US92-09893-4
19	118	3.0	993	5	PCT-US95-03718-4
20	118	3.0	1160	5	PCT-US92-05401-4
21	113	2.9	518	3	US-09-043-123-2
22	110	2.8	30	2	US-08-687-372-6
23	109	2.8	367	3	US-08-467-023-95
24	108	2.8	993	1	US-08-183-211-2
25	108	2.8	993	5	PCT-US95-00176A-2
26	107	2.7	566	2	US-07-862-588B-4
27	107	2.7	920	3	US-08-930-996A-8

28	107	2.7	1248	2	US-08-348-353-17	Sequence 17, Appl
29	107	2.7	1248	2	US-08-465-965-17	Sequence 17, Appl
30	107	2.7	1248	3	US-08-465-966-17	Sequence 17, Appl
31	106.5	2.7	1507	6	5268270-2	Patent No. 5268270
32	105.5	2.7	2048	4	US-09-268-347-48	Sequence 48, Appl
33	105	2.7	1338	2	US-08-728-470-9	Sequence 9, Appl
34	105	2.7	1338	4	US-08-719-641-9	Sequence 9, Appl
35	105	2.7	1599	2	US-08-617-697-9	Sequence 9, Appl
36	104.5	2.7	654	1	US-08-392-828C-2	Sequence 2, Appl
37	104.5	2.7	654	3	US-09-330-945-2	Sequence 2, Appl
38	103.5	2.6	1702	4	US-08-296-791-5	Sequence 5, Appl
39	103.5	2.6	1702	5	PCT-US95-10661A-5	Sequence 5, Appl
40	103	2.6	686	3	US-08-947-965-70	Sequence 70, Appl
41	103	2.6	1222	2	US-08-682-517-15	Sequence 15, Appl
42	103	2.6	1252	2	US-08-682-517-9	Sequence 9, Appl
43	102.5	2.6	1104	4	US-09-268-347-28	Sequence 28, Appl
44	102.5	2.6	1104	4	US-09-268-347-34	Sequence 34, Appl
45	101.5	2.6	591	4	US-09-377-155-11	Sequence 11, Appl

ALIGNMENTS

RESULT 1
US-08-696-944-19
; Sequence 19, Application US/08696944
; Patent No. 5981831
; GENERAL INFORMATION:
; APPLICANT: Sumant CHENGAPPA
; APPLICANT: Susan A. HELLYER
; APPLICANT: John S. REID
; APPLICANT: Jacqueline DE SILVA
; TITLE OF INVENTION: No. 5981831el Exo-(1-4)-Beta-D Galactanase
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pillsbury Madison & Sutro, L.L.P.
; STREET: 1100 New York Avenue, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: MS Word
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/696,944
; FILING DATE: 23-AUG-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/GB95/00372
; FILING DATE: 23-FEB-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9403423.8
; FILING DATE: 23-FEB-1994
; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 838 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-696-944-19

Query Match 74.4%; Score 2920; DB 2; Length 838;
Best Local Similarity 71.6%; Pred. No. 6.6e-254;
Matches 517; Conservative 96; Mismatches 105; Indels 4; Gaps 2;
Qy 1 MLRTNVLILLVLCILDFESSVKASVSDYDRAIIINGKRLISGSIHYPRSTPQWMPDLI 60
Db 6 ILMLNVLVL---LGSVFTASVSDYDRAIIINGKRLISGSIHYPRSTPQWMPDLI 62
Qy 61 QKAKDGLDVIETVFWNGHEPSPCKYNEGRYDLVRFIRKVRAGLYVNLRIQFYVCAE 120

Db 63 QKAKGGVDVDTQYVFWNGHEPQQKYYFEGRYDLVKIKLVHQAGLVXHLRVGPGYACAE 122
Qy 121 WNFEGFPVWLKVPVGMERTNNOPEKVAOGFVOKIVNMKSENLFESQGGPITMAQLEN 180
Db 123 WNFEGFPVWLKVPVGSFRTDNGPKAAQKQFATKIVNMKAERDIETQGGPITLQSLEN 182
Qy 181 EYGPVENEIGAPGKAYTKWAQAQMAVGLTKGVPMWCKQEDAPDPVIDTCNGFYCEGFRPN 240
Db 183 EYGPMEWELGAPGKSYAQAAKMAVGLDTGVPMWCKQDDAPDPPIINACNGFYCDYFSPN 242
Qy 241 KPYKPKMTWETWGTWYKFGGPIQORPAEDIAFSVARVQNNGSFFNYMYHGGTNFGR 300
Db 243 KAYKPKIWEATWATWFTGFGNPVYRPAEDLAFSAVAKIQRGGSFINYMYHGGTNFGR 302
Qy 301 SSGPLIATSYDYDAPLDEYGLNRPKYGHLRDLHKAIKLSEPALVSSYAAVTSLSGNOEA 360
Db 303 AGGPPIATSYDYDAPLDEYGLLRQPKWHLKDLHRAIKLCEPALVSGDPVATLGHQOEA 362
Qy 361 HVIYRSKAGCAAFLSNYSRYSVKVFQNRPNYLPNPPWSISILPDCKTAVYNTAQVNSQSS 420
Db 363 HVIYRSKAGCAAFLANYDOHSFATVSPANRHYNLPNPPWSISILPDCKNVFNARIGAQSA 422
Qy 421 SIKMTPAGGGLSWOSYNEETPTADSDTLTANGWEQKNVTRDSSDYLYWNTVNNIASNE 480
Db 423 QMKMTPTVSRGLPWOSFNEETSSEYEDS-SFTVVGLEQINTTRDVSVDYLYWNTDVKIDSRE 481
Qy 481 GFLKNGKDPYLTVMASAGHLVHFVNGKLSGTVYGTLDNPKLTYSGNKLKRAKINKISLLS 540
Db 482 KFLRGKGPWLTITMSAGHALHFVNGQJAGTAYGSLERPKLTFKAVNLRAGVKNKISLLS 541
Qy 541 VSVGLPNVGVHYDTWNAAGVLPVTLSGLNEGRNLAKOKWSYKVLKGEISLSLSLSSGSS 600
Db 542 IAVGLPNIGPHFETWNAAGVLPVSLTGLDEGRDLTWQKWFYKVLKGEALSLSLSSGSP 601
Qy 601 SVEWVRGSLMAQKQPLTWYKATFNAPGNDPLALDMSMGKQGIWINGEGVGRHWPYVIA 660
Db 602 SVEWVRGSLVAKQKPLSWYKTFNAPDNEPLALDMSMGKQGIWINGEGVGRHWPYVIA 661
Qy 661 QGDCSKCYAGTFNEKKCOTNCGOPQORWYHVPKSWLKPNSGNTLVVFEWGNPTGSLV 720
Db 662 SGCSVCNCTGWFDEKKCLTNCGEGSQRWYHVPKSWLKPNSGNTLVVFEWGNPTGSLV 721
Qy 721 RR 722
Db 722 KR 723

RESULT 2
us-08-696-944-2
; Sequence 2, Application US/08696944
; Patent No. 5981831
; GENERAL INFORMATION:
; APPLICANT: Sumant CHENGAPPA
; APPLICANT: Susan A. HELLIER
; APPLICANT: John S. REID
; APPLICANT: Jacqueline DE SILVA
; TITLE OF INVENTION: No. 5981831el Exo-(1-4)-Beta-D Galactanase
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESS: Pillsbury Madison & Sutro, L.L.P.
; STREET: 1100 New York Avenue, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: MS Word
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/696,944
; FILING DATE: 23-AUG-1996

; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/GB95/00372
; FILING DATE: 23-FEB-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9403423.8
; FILING DATE: 23-FEB-1994
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 730 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-696-944-2

Query Match 73.8%; Score 2896; DB 2; Length 730;
Best Local Similarity 71.4%; Pred. No. 7.6e-252;
Matches 520; Conservative 94; Mismatches 100; Indels 14; Gaps 5;
Qy 1 MLRTN--VLLLV--ICLLDFSSVKASVSYDDRAIINGKRKILISGSIHPRSTPQM 55
Db 12 MSRRNFHMLVLLLFVWCY-----VTASVTIDHAKIMINGQRRILISGSIHPRSTPQM 65
Qy 56 WPDLIQAKAGGLDVIETVYFVWNGHEPSPGKYNEGRDYDLVRFKIMQVORAGLYNLRIQP 115
Db 66 WPDLIQAKAGGLDVIETVYFVWNGHEPSPGKYNEGRDYDLVRFKIMQVORAGLYNLRIQP 125
Qy 116 YVCAEWNFGGPPVWLKVPVGMERTNNOPEKVAOGFVOKIVNMKSENLFESQGGPITM 175
Db 126 FICAEWNGGPPVWLKVPVGMERTNNOPEKVAOGFVOKIVNMKSENLFESQGGPITM 185
Qy 176 AOIENEGVPVWEIGAPGKAYTKWAQAQMAVGLTKGVPMWCKQEDAPDPVIDTCNGFYCE 235
Db 186 SOIENEGVPVWEIGAPGKAYTKWAQAQMAVGLTKGVPMWCKQEDAPDPVIDTCNGFYCE 245
Qy 236 GFRPNKPKMTWETWGTWYKFGGPIQORPAEDIAFSVARVQNNGSFFNYMYHGGT 295
Db 246 NFTPKNYKPLWNTENMTGWYTFAGGATPYRPAEDIAFSVARVQNNGSFFNYMYHGGT 305
Qy 296 NFGTSSGLFTATSYDYDAPLDEYGLNRPKYGHLRDLHKAIKLSEPALVSSYAAVTSLG 355
Db 306 NFGTSSGLFTATSYDYDAPLDEYGLNRPKYGHLRDLHKAIKLSEPALVSSYAAVTSLG 365
Qy 356 SNOEAHVRSKSGACAFLSNYSRYSVKVFQNRPNYLPNPPWSISILPDCKTAVYNTAQV 415
Db 366 KNLEVLHYKTES-ACAAFLANYNTDYSTQVKEGNGQYDLPNPPWSISILPDCKTAVYNTAQV 424
Qy 416 NSQSSSKMTAGGGLSWOSYNEETPTADSDTLTANGWEQKNVTRDSSDYLYWNTVNN 475
Db 425 NSPRLHRRMTVPVNSAFAMQSYNEEPASSSENDDPVTGYALWBEQVGVTRDSSDYLYWNTVNN 484
Qy 476 IASNEGFLKNGKDPYLTVMASAGHLVHFVNGKLSGTVYGTLDNPKLTYSGNVLKRAKINK 535
Db 485 IGPND--IKDGKWPVLTAAMSAGHLVNFVINGQYATAGTAYGSLDDPRLTFESQSVNLRVGNK 542
Qy 536 ISLSVSVGLPNVGVHYDTWNAAGVLPVTLSGLNEGRNLAKOKWSYKVLKGEISLSLS 595
Db 543 ISLSVSVGLANVGTHTETWNTGVLPVTLTGLSSGTWDLKQKWSYKVLKGEISLSLT 602
Qy 596 LSGSSVSEWVRGSLMAQKQPLTWYKATFNAPGNDPLALDMSMGKQGIWINGEGVGRHW 655
Db 603 EAGNSVSEWVRGSLVAKKQPLAWYKTFNAPDNEPLALDMSMGKQGIWINGEGVGRHW 662
Qy 656 PGYTAQGDCKSKCYAGTFNEKKCOTNCGOPQORWYHVPKSWLKPNSGNTLVVFEWGNPT 715
Db 663 PGNKARGNCGNCTAGTYTDTKCLANCGOPQORWYHVPKSWLKPNSGNTLVVFEWGNPT 722
Qy 716 G1SLVRS 723
Db 723 G1ALVERT 730

Query Match	68.2a;	Score	2676;	DB 2;	Length	731;
Best Local Similarity	67.6a;	pred.	No. 4.8e-232;			
Matches	488;	Conservative	89;	Mismatches	137;	Indels
QY	6	VLLLVLLCLDFESSKASYSYDDRAIIINGKRKILISGSIHYPRSTPQMPDLIQAKD	65			
DB	12	MLLVYVFLITLISCVTGNVWYDRAIKINDORILLISGSIHYPRSTPEMMPDLIIENAKD	71			
QY	66	GGLDVIETTYFWNGHGESPKYNEGRIYDLVRFTIKMYQRAGLYVNLRIGPVVCIAENWFGG	125			
DB	72	SQLDVIOQTYVFWNGHGESPKYNEGRIYDLVRFTIKMYQRAGLYVNLRIGPVVCIAENWFGG	131			
QY	126	FPWLKVKVPMGEPRTNQPKVAMQGVOKIVYNNMKSENLFESOGGPIIMAOIENEGPV	185			
DB	132	FPWLKVKVPIEGFTDNGPKQKQVETTKIVDMKMAEKLFHWQGGPIILNQIENEGPV	191			
QY	186	EWEIGAPKAYTTHWAAQMAVGLKTGVPMWCKQE-DAPDPVIDTCNCFYCEGFFPNPKPYK	244			
DB	192	EWEIGAPKAYTTHWAAQMAQSLNAGVPMWCKQSDVDPDNVIDTCNCFYCEGFFPKDKSK	251			
QY	245	PKWTEVWTKWYTKFGGPIQORPAEDTAFSVARVQNNGSFFNYMYHTGGTNFQRTSSGL	304			
DB	252	PKWTEVWTKWYTEYKGPVYRPAEDVAFSVAREIQNGSFFMYMYHTGGTNF-ETTAGR	310			
QY	305	FIATSYDYDAPLDEYGLNLNPKYGHRLDKLHKAIKLSEPALVSSVAAVYTLGSGNQEAHYR	364			
DB	311	FVSTSYDYDAPLDEYGLPREPKYTHLKNLHKAIKMCEPALVSSDAKYTNLGSNQEAHYS	370			
QY	365	SKSGACAAPLISNDSRYSVKVTFQNRPNYLPPWISITLPOCKTAYVNTQVNSOSSSI--	422			
DB	371	SNSGACAAPLANYDPKWSKVYTFSGMEFELPAWISITLPOCKKQVYARTARNEPSKLHS	430			

[illegible]

Db 258 KSEPKGLVNSBYTCGLDHWKOPHSTVTEVVASSL-HDIIAAGANVNLWYFVIGTNEFA 316
QY 298 ---GRTSSGLFIATSYDYDAPLDEYGLLNEPKYGHRLDLHKA-----IKLSEPALVS 346
Db 317 YWNGANMPYQAOPTSYDYDAPLSEAGDLTE-KYFALREVIRKEFKVPEGFIPSPFKFAY 375
QY 347 SYAATVSLGSNOBAHYRYSKSGACAAFLSNYSRYSVKVTFQNRPNYPLPWSISILPDCK 406
Db 376 GRVALKLLKTBEEALNVLCPGPG-----INSLYPLTEIQVKQVGFVWYRTTLPDQCS 428
QY 407 TAVYNTAQVNSOSSSTKMPAGGLSWQSYNEETPTADDSDLTANGLWEQKNVTRDSD 466
Db 429 DP-----TPLSSPLS-----GVHDRAVSVVG-- 450
QY 467 YLWYMTNVNIASNEGFKNKGKDPYLTVMAGHVLHVFNKLSGTGYGLDNPCKLTYSGN 526
Db 451 -----VPOG-----VMERSNVITLITGKAGAT----- 473
QY 527 VKLRAGINKISLSVGLPNVGVHYDTWNAGVLPVTLISGLNEGSRLNAKOKW----- 580
Db 474 -----LDLLVENMGRVNYGRYINDFK-GLISNLT-----GSSILT--NMWIFPLN 516
QY 581 -----SYKVLKLGESLSLH--SLGSSSVVWVRSGLMAQKQPLTWYKATFNAGGNDPL 632
Db 517 TEDAVRSHLGGWHGPNNGRHRDXTFAHRSS-----NYTLPA-----FYMGNFSIPSGIDPL 566
QY 633 ALD-----MASMGKGQIWIINGEGVGRHWP 656
Db 567 PDDTFIQPPGWTKGQWINGFNLGRYWP 594

RESULT 5

US-08-596-985-2
; Sequence 2, Application US/08596985
; Patent No. 5736374
; GENERAL INFORMATION:
; APPLICANT: Berke, Randy M.
; APPLICANT: Hucul, John A.
; APPLICANT: Ward, Michael
; TITLE OF INVENTION: Increased Production of
; TITLE OF INVENTION: Beta-galactosidase in Aspergillus oryzae
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESS: Genencor International, Inc
; STREET: 180 Kimball Way
; CITY: South San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/596.985
; FILING DATE: 05-FEB-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/267,631
; FILING DATE: 29-JUN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Horn, Margaret A
; REGISTRATION NUMBER: 33,401
; REFERENCE/DOCKET NUMBER: GC250
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 742-7536
; TELEFAX: (415) 742-7217
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 911 amino acids
; TYPE: amino acid
; STRANDEDNESS: single

; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-596-985-2
Query Match 10.1%; Score 396.5; DB 1; Length 911;
Best Local Similarity 23.0%; Pred. No. 1.1e-26;
Matches 197; Conservative 98; Mismatches 316; Indels 245; Gaps 36;
QY 25 VSYDDRAITINGKRKTLISGSIH-YPRSTPQMPDLOIAKADGGDLVDIETVYFWNGHEPS 83
Db 46 VTWDDKSLFINGERIMLFSGEVHPFLPVLPSLWLDIFKIRALGFCVCFYIDWALLEK 105
QY 84 PGKYFEGRYDLVRFIKMQRAGLYVNLGRIGYPVCAEWNFGGPPVWLKYVPGNEFTNNQ 143
Db 106 PGDYRAEGIFALEPFPDAKAAGIIVLIARPGSYINAEVSGGPGWLQVRNG-TLRSSDE 164
QY 144 PKFVAMQGVQKIVNMKSENLFESQGGPIIMAOITENEYGPVWEIGAPKAYT-----K 198
Db 165 PFLKATDNYIANAAAAVAKAQI--TNGGPIVLYQENYSG-----GCCGVKYTADYMQ 217
QY 199 WAAQNAVGLKTGPVIMCKQEDAPD-----PVIDTCN-----GFYC-----EG 236
Db 218 YWDOARKADIIVPFI--SNDASPSGHNAPGSGTGAVDIYGHDSPLGDCANPSWVPEG 275
QY 237 FRPN-----KPKPKMTTEVMTGWTKFGGPIQRPAPEDIAFVSAR-EVQNNGSP- 285
Db 276 KLDPNFTLHLQSPSAPYSLLEFOAGAFDPWGGPGFEKCYALVNHFSRVFYRNDLSFG 335
QY 286 ---FNYMYHGTNFGRTS-SGLFIATSYDYDAPLDEYGLLNEPKYGHRLDLHKAIKLSE 341
Db 336 VSTFNLMTFGTNGNLGHPGGY--TSYDYGSPITETRNVTREKYSIDIKLANEFKASP 393
QY 342 PALV-----SSYAAVTSI-----CSNOEAHVYRSKSGACAA- 372
Db 394 SYLTATPRLNLTGVTYDTSDLAVTPLIGDSPGSFFVVRHTDYSSQSESTYKLUKPTSAGN 453
QY 373 -----FLSNYDSRYSVKVTQNRPNYLPWPSISILPCKTAVYNTAQNSSSI 422
Db 454 LTIPQLEGTLSLNGRDSKTHVV-----DYNVSGTNI-----IYSTAEV----- 491
QY 423 KMPAGGGLSWQSY-----NEETPTADDSLTTLTANGLWEQKNVTRDSDYL 468
Db 492 -----FTWKKFDGNKVLVLYGGPKHEHETAIASKSNVTIIEGSDSIVSTRKSSVI 543
QY 469 --WYMTNVNIASNEGFKNKGKDPYLTVMAGHVLHVFNKLSGTGYGLDNPCKLTYSGN 526
Db 544 IGV-----DVSSTRIVQVGD-----LRVFLLGKNSAYNYWVPELP--TEGTS 584
QY 527 VKLRAGINKISLSVGLPNVGVHYD-----TNAGVLPVTLISGLNEGSRLN----- 575
Db 585 PGFSTSKTTASSIIVKAGYLLRGALHDGADLHLTADFNATTPTEVIGATGAKNLFVNGE 644
QY 576 -----AKQWSYKVLKGESLSLHSL-----SCSSSVWVWVRSGLMAQK 613
Db 645 KASHTVDXNGIWSSEVKYAAPEIKPLGLKDLDMKLDLTLPETIKSYDDSAWYSADLPKTK 704
QY 614 ---QPLTW-----YKATFNAPGNDPLAL--DMASMGKGQIWIINGEV 651
Db 705 NTHRPDLTPTSLYSSDYGFHTGYLIYRGHFVANGKESEFLINTQGSAGFSSVWLNEYTL 764
QY 652 GRHWPYIAQGDCKSKCYAGTFNEKKCQTNCGQPSQRWYH-----PRSWLKPSG 701
Db 765 G-SWTG-----ADYTMDSNSTYKLSQLESN-----YHVTIVIDLNLGLDENW----- 806
QY 702 NLLVVFEEWGNPTGI 717
Db 807 ---TVGEETMKNPRGI 819

RESULT 6

US-08-551-459-4
; Sequence 4, Application US/08551459

Patent No. 5821350
; GENERAL INFORMATION:
; APPLICANT: Huang, Yue
; APPLICANT: Karatzas, Costas N.
; APPLICANT: Lazaris-Karatzas, Anthoula
; APPLICANT: Delaquis, Annick
; TITLE OF INVENTION: ASPERGILLUS NIGER BETA GALACTOSIDASE
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/551,459
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Clark, Paul T.
; REGISTRATION NUMBER: 30,162
; REFERENCE/DOCKET NUMBER: 06632/005001
; TELEPHONE: 617/542-5070
; TELEFAX: 617/542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1007 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-551-459-4

Query Match 10.1%; Score 395; DB 2; Length 1007;
Best Local Similarity 23.4%; Pred. No. 1.8e-26;
Matches 187; Conservative 103; Mismatches 308; Indels 200; Gaps 32;
Qy 25 VSYDDRAIIINGKRIISGSIHPRSTP--QMPPDLQKAKDGLDVIETVFWNGHEP 82
Db 46 VTWDDKSLFINGERIMFSGEFH-PFRLPVKELQDIFQKVKALGFNCVSYVDWALVEG 104
Qy 83 SPGKYNFEGDYDLVRFKMWQAGLYVNLRIQPVYCAEWNFGGFPVWLKYVPGMFEFTNN 142
Db 105 KPGEYRAGDIFDLFPFPAASEAGIYLARPGYINAESSGGGFPGLQRYNG-TLRSSD 163
Qy 143 QPFVKVMOGFQKIVNMKMSLENFESQGGPIIMAEIENYGPVWEIGAPGKAYTKWAAQ 202
Db 164 KAYLDATDNVYSHVAATIAKYOI--TNGGPITLYOPENEYSGCGVEFPDPVYMQYVED 221
Qy 203 MAVGLKTGVPIH---MCKQEDAPP---VIDTCN-----GFYC-----EGF 237
Db 222 QARNAGVVIPLINDASASGNAPGTGKGAVIDYGHDSYPLGFCDCANFTVWPSGLPTNF 281
Qy 238 R----PNKPYKPKMTEVMTGWYTKFGGPIQRPADIEDAFVAR-FVONNGSF---FNY 288
Db 282 RTLHLEQSPPTPYAIVEFGGSDYDWPWGGPGGAACSELNINFEFVYKRNDFSFQTAIWNL 341
Qy 289 YMHGGTNFGRTS---SGLFIATSYDYDAPLDEYGLLNEPKYGHRLDLHLKAIKLSPALVS 346
Db 342 YMIFGTTWNGNLGYPNGY---TSYDYGSAVTESTRNITREKYSSELKLLGNFAKVPSPGYLTA 398
Qy 347 S-----YAAVTSLG-----SQEAAHVYRSKSGACAA----- 372
Db 399 SPGNLTSGYADTTDLTVTLPLGNSTGSGFFVVRHSDYSSESTSYKLRPLTSAGSVTIPO 458

Qy 373 -----FLSNYDSRYSVKVTQNRPNLPPWSISILPDCKTAVYNTAQV----- 415
Db 459 LGCTITLNGRDSKIHV-----TDYVSGTNI-----IYSTAEVFTWKKFADGKVL 503
Qy 416 -----NSQSSIRKMTFAGGGLSWQSYNEETPTADSDT-----LTFANGLWEOK 458
Db 504 VLYGGAGEHHELAISTKSNVTIEGSESGISKQTSSSVVWGVWDVSTTRRIIOVGDLKIL 563
Qy 459 NVTROSSDYLYWYMTNVNNTASNEGLKNGKDPYLYTMSAGHVLHVFNGLSGTIVYCTLDN 518
Db 564 LLDRNSAYNWVWPOLATDGTSPGFSTPEKVASSIIVKAGYLVR-----TAY----- 609
Qy 519 PKLTYSGNVKLRAGINKISLISVSGLPNVGVHYDTWAGVLGVPVTLISLNGSGSRNLAKQ 578
Db 610 --LKSGG-LYLTADFNAATSVVEV-IGVPSTAK-----NLFINGDKTSHVDKNG----- 654
Qy 579 KWSYKVGKLGESLSLSLSG-----SSSVE---WVRGSLMAKQKPL----- 616
Db 655 IWSATVDYNAPDISLSPLKDLWKYVDTLPEITQSSYDDSLWPAADLKQTKNLTLSLTPT 714
Qy 617 -----TWYKATFNAPGGNDPLALDM--ASMCKGOIWIINGEGVGRHWPYIAQ 661
Db 715 SLYSSDYGFHTGYLLYRGHFTATGNESTFAIDTQGSAGFSSVWLNGLTYLG-SWTGLYAN 773
Qy 662 GDCSKCSYAGTFENEKKCQ 679
Db 774 SD-----YNATYNLPQLQ 786

RESULT 7
US-08-840-062-5
; Sequence 5, Application US/08840062
; Patent No. 6117977
; GENERAL INFORMATION:
; APPLICANT: LASKY, LAURENCE A.
; TITLE OF INVENTION: TYPE C LECTINS
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WinPatIn (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/840,062
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Dreger, Ginger R.
; REGISTRATION NUMBER: 33,055
; REFERENCE/DOCKET NUMBER: P1019R1
; TELEPHONE: 415/225-3216
; TELEFAX: 415/952-9881
; TELEX: 910/371-7168
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1455 amino acids
; TYPE: Amino Acid
; TOPOLOGY: Linear
US-08-840-062-5

Query Match 3.7%; Score 146.5; DB 3; Length 1455;
Best Local Similarity 20.1%; Pred. No. 0.00076;

```

; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/687,372
; FILING DATE: 2-AUG-1996
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9320930.2
; FILING DATE: 12-OCT-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/GB94/02203
; FILING DATE: 10-OCT-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: KOKULIS, PAUL N.
; REGISTRATION NUMBER: 16,773
; REFERENCE/DOCKET NUMBER: 224239/SEE37803X/UST
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 861-3000
; TELEFAX: (202) 822-0944
; TELEX: 6714627 CUSH
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 30 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-687-372-1

```

Query Match 3.6%; Score 141; DB 2; Length 30;
 Best Local Similarity 93.3%; Pred.No. 3.8e-06;
 Matches 28; Conservative 0; Mismatches 2; Indels

```

QY 24 SVSYDDRAIIINGKRKILISGSIHYPRSTP 53
      |||||
DB 1 SVSYDDRAIIINGKRKILISGSIHYPRKYP 30

```

```

RESULT 9
US-08-687-372-7
; Sequence 7, Application US/08687372
; Patent No. 5859344
; GENERAL INFORMATION:
; APPLICANT: BIRD, COLIN R
; APPLICANT: HOLT, KAREN A
; APPLICANT: PICARD, SYLVIE

```

APPLICANT: CAREY, ANNETTE I
APPLICANT: SEYMOUR, GRAHAM B
APPLICANT: TUCKER, GREGORY A.
TITLE OF INVENTION: MODIFIED FRUIT
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: CUSHMAN DARBY & CUSHMAN, L.L.P.
STREET: 1100 New York Avenue, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20005-3918
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/687,372
FILING DATE: 2-AUG-1996
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9320930.2

```
; FILING DATE: 12-OCT-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/GB94/02203
; FILING DATE: 10-OCT-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: KOKULIS, PAUL N.
; REGISTRATION NUMBER: 16,773
; REFERENCE/DOCKET NUMBER: 224239/SEE37803X/UST
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 861-3000
; TELEFAX: (202) 822-0944
; TELEX: 6714627 CUSH
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 30 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-687-372-7

Query Match          3.5%; Score 137; DB 2; Length 30;
Best Local Similarity 86.7%; Pred. No. 8.7e-06;
Matches 26; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 24 SVSYDRAIIINGKRKILISGSIHYPRSTP 53
DB 1 SVSYDHKALIIVGQRKILISGSIHYPRSTP 30

RESULT 10
US-08-467-023-2
; Sequence 2, Application US/08467023
; Patent No. 6090386
; GENERAL INFORMATION:
; APPLICANT: Griffith, Irwin J.;
; APPLICANT: Pollock, Joanne;
; APPLICANT: Bond, Julian F.;
; APPLICANT: Garman, Richard D.;
; APPLICANT: Kuo, Mei-Chang;
; APPLICANT: Yeung, Siu-mei H.;
; APPLICANT: Brauer, Andrew;
; APPLICANT: Exley, Mark A.;
; APPLICANT: Powers, Steven P.
; TITLE OF INVENTION: Allergenic Proteins And Peptides From
; TITLE OF INVENTION: Japanese Cedar Pollen
; NUMBER OF SEQUENCES: 261
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Immunologic Pharmaceutical Corporation, Inc.
; STREET: 610 Lincoln St
; CITY: Waltham
; STATE: MA
; COUNTRY: USA
; ZIP: 02154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/467.023
; FILING DATE: June 6, 1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/350,225
; FILING DATE: December 6, 1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Jane E. Remillard
; REGISTRATION NUMBER: 38,872
; REFERENCE/DOCKET NUMBER: 025.6 USD2 (IMI-028CPD2)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 227-7400
```

```
; TELEFAX: (617) 227-5941
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 374 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-467-023-2

Query Match          3.2%; Score 125; DB 3; Length 374;
Best Local Similarity 21.7%; Pred. No. 0.0069;
Matches 89; Conservative 38; Mismatches 139; Indels 144; Gaps 21;

QY 276 ARFVONGSFFPNVYMHGTFNFGRTSSGLFIATSYDYDAPLDDEYGLNPKYGHRLDL-- 333
DB 33 SNWAQRNKLADCAVGFSGSSTMGKGDDLYTNTSDDDPVNPAPGFL---RYGATRDRL 89
QY 334 -----HKAIKLSEPALVSSYAAVTSLSGNOEAAHYVYRSKGACAAFLSNYSRYSVKVTF 387
DB 90 WIIFSGNMNIKLMKPMYIAGYKTFDGRG---AQVYIGNGGPC-----VF 130
QY 388 QNRPNLPPWSISILPDCKTAVYNTAQVNSQSSSIKMTFAGGGLSWQSYNEETPTADSD 447
DB 131 IKRVSNVIHGL-YLYGCTSVLGNVLN-----ESFGVEPVHPQDGD 172
QY 448 TL---TANGLWEQKNVTRDSDYLWYMTNVTNIA---SNEGFLKNGKDPYLTVMASAGHVLH 501
DB 173 ALTLRTAINIWDHNSFSNSDGLVDVTLTSTGVTISNNLFFNHHK-----VLLGH--- 224
QY 502 VEVNGKLSGTVYGTLDNPKLTYSG--NVKLRAGINKISLLSVSLPNVG----- 549
DB 225 -----DD---AYSDDKSMKVTVAFNQFG-----PNCQRMPRARYGL 258
QY 550 VH-----YDTWAGVLG-----PVTLSGLNEGRNLAKQKWSYKVLKGBESLSLSLSSGS 600
DB 259 VHVANNYDPWTIYAIGGSSNPTILS---EGNSFTAPNE-SYK---KQVTIRIGCKTSS 311
QY 601 SVEWV-----RGSMAOKOPLTWYKATFNAPGN 629
DB 312 CSNWWQSTQDYFYNGAYFVSSGKYEGGNIYTKKE-----AFNVENG 354

RESULT 11
US-07-977-451-4
; Sequence 4, Application US/07977451
; Patent No. 5270458
; GENERAL INFORMATION:
; APPLICANT: Lemischka, Ihor R.
; TITLE OF INVENTION: TOTIPOTENT HEMATOPOIETIC STEM CELL
; TITLE OF INVENTION: RECEPTORS AND THEIR LIGANDS
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Imclone Systems Incorporated
; STREET: 180 Varlick Street
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10014
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/977.451
; FILING DATE: 19921119
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US UNASSIGNED
; FILING DATE: 12-NOV-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/906,397
```

```

; FILING DATE: 26-JUN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US PCT/US92/05401
; FILING DATE: 26-JUN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: TW 81102961
; FILING DATE: 15-APR-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US PCT/US92/02750
; FILING DATE: 02-APR-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/813,593
; FILING DATE: 24-DEC-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/793,065
; FILING DATE: 15-NOV-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/728,913
; FILING DATE: 28-JUN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/679,666
; FILING DATE: 02-APR-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Feit, Irving N.
; REGISTRATION NUMBER: 28,601
; REFERENCE/DOCKET NUMBER: LEM-3-7P
; TELEPHONE: 212-645-2054
; TELEFAX: 212-645-2054
; INFORMATION FOR SEQ ID NO: 4:
; LENGTH: 993 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-07-977-451-4

Query Match 3.0%; Score 118; DB 1; Length 993;
Best Local Similarity 18.5%; Pred. No. 0.15;
Matches 111; Conservative 74; Mismatches 199; Indels 216; Gaps 30;

QY 61 QKAKGGGLDVITYVFNWGHSPGKYNFEGYDLVRFKTMQVORAGLYVNLRIQPVVCAE 120
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 394 QKGLDNGYSISK---PCN-HKHQPGGEYIFHAENDDAQFTKM-----FTLNIRKKPOVLAE 444
QY 121 WNFQ-----GFPVWLKVPVPGMEFTNQPCKVAMQGFQVKIVNMKSENLFESOGPI 173
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 445 ASASQSCFSDGYP-----LPSWTWKKCD-----KSPNCTE----- 476
QY 174 IMAQIENYGPVWEITGAPGKAYTKWAQMAVGLKTGVPWIMCKQEDAPDPVIDTCN--- 230
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 477 -----EITEGVNRKANRKGQVGVSSSTLNMSAIGKFLVK-----CCAYNS 519
QY 231 -GFCGGRPNPKYPKPMWTEWTGWTYKFGGPIQORPAEDIAFSVARVQNGSFF--- 286
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 520 LCTSCETILLNSP-----GPPF-----FIQDNISFYATI 548
QY 287 -----NYYMHGCTNF-----GRTSSGLFIA--TSYDYDAPLD--- 317
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 549 GVCLLFIVVTLTLLCHYKKQRYESQLOMQVQVTGSSDNEYFYDFREYEDLKWEFPRE 608
QY 318 --EYG-----LLNEPKYGHLR---DLHKAIKL-----SEPALVSSYAAVTSL 354
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 609 NLEFGKVLGSGAFGKVMNATAGISKTVSIOVAVKMLKEKADSSREALMSELKMMTOL 668
QY 355 GSNQ-----EAHYVRKSGACAFLSNY-----DSRYSVKVT 386
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 669 GSHENIVNLLGACTLGGPIYLIFEYCCYDGLLNLRSKREKFRHTWTETFEKHNFSEYPT 728
QY 387 FQNRPNLPWISILPDCKTAVYNTAQVNSQSSSIKMTWPAGGGLSQWSQYNEETPADDS 446
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 729 FQSHPNSSMPGSGREV-----QIHPDSQDI-----SGLHGNFSHSE---DEI 766
```

```

QY 447 DLTANGLWEQKNV-TRDSSDYLWYMTNNTNVIASNEGFLKNGKDPYLTVMAGHLVHFVN 505
; : | : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 767 EYENQKRLSEEDNLVLTFTEDLLCFA--YQVAKGMEFLE-FKSCVHRDLAARNVL--VTH 821
QY 506 GKL-----SGTVYGTLDNPKLTYSGNVKLRAGINKISLISVSVGLPNVGVHYDTWAGVL 560
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 822 GKVVKICDFGLARDIMSDSNYYVVRGNARL--PVKWMAPESLFEGI--YTIKSDVWSYGIL 877

RESULT 12
US-08-252-517-4
; Sequence 4, Application US/08252517
; Patent No. 5548065
; GENERAL INFORMATION:
; APPLICANT: Lemischka, Ihor R.
; TITLE OF INVENTION: TOTIPOTENT HEMATOPOIETIC STEM CELL
; TITLE OF INVENTION: RECEPTORS AND THEIR LIGANDS
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ImClone Systems Incorporated
; STREET: 180 Varick Street
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10014
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/252,517
; FILING DATE: 31-OCT-1994
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/977,451
; FILING DATE: 19-NOV-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/906,397
; FILING DATE: 26-JUN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US PCT/US92/05401
; FILING DATE: 26-JUN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: TW 81102961
; FILING DATE: 15-APR-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US PCT/US92/02750
; FILING DATE: 02-APR-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/813,593
; FILING DATE: 24-DEC-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/793,065
; FILING DATE: 15-NOV-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/728,913
; FILING DATE: 28-JUN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/679,666
; FILING DATE: 02-APR-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Feit, Irving N.
; REGISTRATION NUMBER: 28,601
; REFERENCE/DOCKET NUMBER: LEM-3-7P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-645-1405
; TELEFAX: 212-645-2054
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 993 amino acids
; TYPE: amino acid
```

TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-252-517-4

Query Match 3.0%; Score 118; DB 1; Length 993;
Best Local Similarity 18.5%; Pred. No. 0.15;
Matches 111; Conservative 74; Mismatches 199; Indels 216; Gaps 30;

Qy 61 QKADGGDLVETVYFVWNGHEPSPGKYNFEGRYDLVRFIKMKVQRAGLVNLRIQYVCAE 120
Db 394 QKGLDNGYSISK---FCN-HKHQGEYIFHAENDDAQFTKM-----FTLNIRRKQVLA 444
Qy 121 WNFQ-----GFPVWLKYVPCMEFRTNNOPFKVAMQGVQKIVNMKMSLENFESQGGPI 173
Db 445 ASASQASCFSDGYP-----LPSWTWKCCSD-----KSPNCTE----- 476
Qy 174 IMAQIENYGPVWEIGAPGKAYTKWAQMAVGLKTGVPMCKOEDADPDVIDTCN--- 230
Db 477 -----EITEGVNMRKANRVFGQWSSSTLNMSEAIKGLVK-----CCAYNS 519
Qy 231 -GFYCEGFRPNKPKMKMTWVTGWTGKFGPIQORPAEDIAFSVAREVQNGSFF--- 286
Db 520 LGTSCETILLNSP-----GPPF-----FIQDNISFYATI 548
Qy 287 -----NYMYHGGTNF-----GRTSSGLFIA--TSYDYDAPLD--- 317
Db 549 GVCLLFIVVLTLLICHYKQFRYESQLQMVQVTCSSDNEYFYVDFREYEDLKWEFPRE 608
Qy 318 --EYG-----LLNEPKYGHLR---DLHKAIKL-----SEPALVSSYAAVTS 354
Db 609 NLEFGKVLGSGAFGKVMNATAYGSKTGVSIOQAVKMLKEKADSSEREALMSKMTQL 668
Qy 355 GSNQ-----EAHVYRSKSGACAFLSNY-----DSRYSVKVT 386
Db 669 GSHENIVNLLGACTLSGPIYLIFECYCGDLLNLYRSKREKFRHTWTEIFKEHNF 728
Qy 387 FQNRPNYLPWWSISITLPDCKTAVYNTAQNVSQSSIKMTPAGGGLSWQSYNEETPTADDS 446
Db 729 FQSHPNSSMPGSGREV-----QIHPDSQI-----SGLHGNSFHSE---DEI 766
Qy 447 DTLTANGLWEQKNV--TRDSSDYLWYMTNVIASNEGFLKNGKDPYLTVMASAGHVLHVFVN 505
Db 767 EYENOKRLEEEEDNLVTFEDLLCFA--YQVAKMGFLE-FKSCVHRDLAARNVL--VTH 821
Qy 506 GKL-----SGTVYGTLDNPKLTYSQNVKLRAGINKISLLSVSGLPNVGVHYDTWNA 560
Db 822 GKVKICDFGLARDIMSDSNVYVRGNARL--PVKWMAPESLPEGI--YTIKSDVWSYGIL 877

RESULT 13
US-07-906-397A-4
Sequence 4, Application US/07906397A
Patent No. 5621090
GENERAL INFORMATION:
APPLICANT: Lemischka, Ihor R.
TITLE OF INVENTION: TOTIPOTENT HEMATOPOIETIC STEM CELL
TITLE OF INVENTION: RECEPTORS AND THEIR LIGANDS
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: INCLONE SYSTEMS INCORPORATED
STREET: 180 VARICK STREET
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: U.S.A.
ZIP: 10014
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/906.397A

FILING DATE: 19920626
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/813,593
FILING DATE: 24-DEC-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/793,065
FILING DATE: 15-NOV-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/728,913
FILING DATE: 28-JUN-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/679,666
FILING DATE: 02-APR-1991
ATTORNEY/AGENT INFORMATION:
NAME: Feit, Irving N.
REGISTRATION NUMBER: 28,601
REFERENCE/DOCKET NUMBER: LEM-3-PPPPPP
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-645-1405
TELEFAX: 212-645-2054
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 993 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-906-397A-4

Query Match 3.0%; Score 118; DB 1; Length 993;
Best Local Similarity 18.5%; Pred. No. 0.15;
Matches 111; Conservative 74; Mismatches 199; Indels 216; Gaps 30;
Qy 61 QKADGGDLVETVYFVWNGHEPSPGKYNFEGRYDLVRFIKMKVQRAGLVNLRIQYVCAE 120
Db 394 QKGLDNGYSISK---FCN-HKHQGEYIFHAENDDAQFTKM-----FTLNIRRKQVLA 444
Qy 121 WNFQ-----GFPVWLKYVPCMEFRTNNOPFKVAMQGVQKIVNMKMSLENFESQGGPI 173
Db 445 ASASQASCFSDGYP-----LPSWTWKCCSD-----KSPNCTE----- 476
Qy 174 IMAQIENYGPVWEIGAPGKAYTKWAQMAVGLKTGVPMCKOEDADPDVIDTCN--- 230
Db 477 -----EITEGVNMRKANRVFGQWSSSTLNMSEAIKGLVK-----CCAYNS 519
Qy 231 -GFYCEGFRPNKPKMKMTWVTGWTGKFGPIQORPAEDIAFSVAREVQNGSFF--- 286
Db 520 LGTSCETILLNSP-----GPPF-----FIQDNISFYATI 548
Qy 287 -----NYMYHGGTNF-----GRTSSGLFIA--TSYDYDAPLD--- 317
Db 549 GVCLLFIVVLTLLICHYKQFRYESQLQMVQVTCSSDNEYFYVDFREYEDLKWEFPRE 608
Qy 318 --EYG-----LLNEPKYGHLR---DLHKAIKL-----SEPALVSSYAAVTS 354
Db 609 NLEFGKVLGSGAFGKVMNATAYGSKTGVSIOQAVKMLKEKADSSEREALMSKMTQL 668
Qy 355 GSNQ-----EAHVYRSKSGACAFLSNY-----DSRYSVKVT 386
Db 669 GSHENIVNLLGACTLSGPIYLIFECYCGDLLNLYRSKREKFRHTWTEIFKEHNF 728
Qy 387 FQNRPNYLPWWSISITLPDCKTAVYNTAQNVSQSSIKMTPAGGGLSWQSYNEETPTADDS 446
Db 729 FQSHPNSSMPGSGREV-----QIHPDSQI-----SGLHGNSFHSE---DEI 766
Qy 447 DTLTANGLWEQKNV--TRDSSDYLWYMTNVIASNEGFLKNGKDPYLTVMASAGHVLHVFVN 505
Db 767 EYENOKRLEEEEDNLVTFEDLLCFA--YQVAKMGFLE-FKSCVHRDLAARNVL--VTH 821
Qy 506 GKL-----SGTVYGTLDNPKLTYSQNVKLRAGINKISLLSVSGLPNVGVHYDTWNA 560
Db 822 GKVKICDFGLARDIMSDSNVYVRGNARL--PVKWMAPESLPEGI--YTIKSDVWSYGIL 877

RESULT 14
US-08-222-299-4
; Sequence 4, Application US/08222299
; Patent No. 5635388
; GENERAL INFORMATION:
; APPLICANT: Bennett, Brian D.
; APPLICANT: Broz, Susan D.
; APPLICANT: Matthews, William
; APPLICANT: Ziegler, Francis C.
; TITLE OF INVENTION: AGONIST ANTIBODIES AGAINST THE FLK2/FLT3 RECEPTOR AND USES THEREOF
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: patin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/222,299
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Hasak, Janet E.
; REGISTRATION NUMBER: 28,616
; REFERENCE/DOCKET NUMBER: 879
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/225-1896
; TELEFAX: 415/952-9881
; TELEX: 910/371-7168
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 993 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; US-08-222-299-4

Query Match	3.0%	Score 118;	DB 1;	Length 993;
Best Local Similarity	18.5%;	Pred. No. 0.15;		
Matches 111; Conservative	74;	Mismatches 199;	Indels 216;	Gaps 30;
QY	61	QKAKDGGLDVETVFNNGHEPSPGKNFYEGRYDLVRFIKMKVQRAGLYVNLRIGPYVCAE	120	
Db	394	OKGLDNGYSISK---FCN-KHKOPEGXYIFHAENDDAQFTKM-----FTLNIRRRPOVLAE	444	
QY	121	WNFG-----GRPVWLKYVPGMFRFNNOPFKVAMOGFVKIVNMKMSLENFSQGSGPI	173	
Db	445	ASASQASCFSDBGYP----LPSWTWKKCSD-----KSPNCTE-----	476	
QY	174	IMAQIENEYGFVEWEIGAPQKAYTKWAAQMAVGKLTGPVWIMCKQEDAPDVITCN---	230	
Db	477	-----EITEGVWNRKANRKKVFQGWSSSFLNNSEAIKGLFVK-----CCAVNS	519	
QY	231	-GFVCBGERNKPKYPMKWTEVTGWTKTGGPTIPQRPEDIAESVARFVONNGSFP---	286	
Db	520	LGTSCETILLNSP-----GPPP-----FIODNISFYATI	548	
QY	287	-----NYMYHGGTGTF---GRSSSGLFIA--TSYDYDADPLD----	317	
Db	549	GVCLLFITWLLTLLCHYKKQFRYESQLQMVOVQTGSSDSNEYFYDFREYEDLKLWEPRE	608	
QY	318	--EYG-----LLNEPKYIGHLR--DIJHKAIKL-----SRPALVSSYAATSL	354	

```

1:1          : 1:1          : 1:1          : 1:1          : 1:1          : 1:1          : 1:1          : 1:1          :
609 NLEPGKVLGSGARGKVMNATYIGISKTGYIQVAVKMLKEKADSSEREALMSLKMWTOL 666
QY 355 GSNO-----EAHYVYKSGACAAFLSNY-----DSRYSVKVT 386
1:1          : 1:1          : 1:1          : 1:1          : 1:1          : 1:1          : 1:1          : 1:1          :
669 GSHENIVNLLGACTLSGPILVIFEYCCYGLLNLRSKREKFHRTWTEIFKHEHNFSPYPT 728
QY 387 FQNPYNLPPWSTISLPDCKTAYNTAQNVSQSSSIKMTFAGGLSWQSYNEETPTADD 446
1:1          : 1:1          : 1:1          : 1:1          : 1:1          : 1:1          : 1:1          : 1:1          :
729 FQSPHNSMPGSRREV-----QHPDSDQI-----SGLHGNFSHSE----DEI 766
QY 447 DTLTANGLWEPQKNV--TRDSSDYLMWMTNVNIASNEGFLKNGKDPYLTVMASAGHVLHFVN 505
1:1          : 1:1          : 1:1          : 1:1          : 1:1          : 1:1          : 1:1          : 1:1          :
767 EYENQKLESEEDLNVLTFEDLLCFA--YQVAKGMEFLE-FKSCVHRDLAARNVL--VTH 821
QY 506 GKL-----STVYGTLDNPKLTYSNNKURAGINKISLISVSGLPNVGHHYDTWNA 560
1:1          : 1:1          : 1:1          : 1:1          : 1:1          : 1:1          : 1:1          : 1:1          :
822 GKVVYKICDFGLARDIMSDSNVYVRGNARL--PVKWMAPESLFEGI--YTIKSDVM 877
SYGIL

RESULT 15
US-08-601-891-4
; Sequence 4, Application US/08601891
; Patent No. 5747651
; GENERAL INFORMATION:
; APPLICANT: Lemischka, Ihor R.
; TITLE OF INVENTION: TOTIPORENT HEMATOPOIETIC STEM CELL
; TITLE OF INVENTION: RECEPTORS AND THEIR LIGANDS
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ImClone Systems Incorporated
; STREET: 180 Varick Street
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10014
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/601,891
; FILING DATE: 15-FEB-1996
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/977,451
; FILING DATE: 19-NOV-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/906,397
; FILING DATE: 26-JUN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US PCT/US92/05401
; FILING DATE: 26-JUN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: TW 81102961
; FILING DATE: 15-APR-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US PCT/US92/02750
; FILING DATE: 02-APR-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/813,593
; FILING DATE: 24-DEC-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/793,065
; FILING DATE: 15-NOV-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/728,913
; FILING DATE: 28-JUN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/679,666
; FILING DATE: 02-APR-1991
; ATTORNEY/AGENT INFORMATION:

```



```
; NAME: Feit, Irving N.
; REGISTRATION NUMBER: 28,601
; REFERENCE/DOCKET NUMBER: LEM-3-7P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-645-1405
; TELEFAX: 212-645-2054
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 993 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-601-891-4

Query Match      3.0%; Score 118; DB 1; Length 993;
Best Local Similarity 18.5%; Pred. No. 0.15;
Matches 111; Conservative 74; Mismatches 199; Indels 216; Gaps 30;

QY 61 QKAKDGLDVIETVFWNGHEPSPGKYNFEGRYDLVRFIKMVORAGLYVNLRIQPVYCAE 120
Db   ||| : : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
394 QKGLDNGYSISK---FCN-HKHQGEYIFHAENDDAQFTKM-----FTLNIRKPOVLAE 444
QY 121 WNEG-----GPPVWLKYVPGHEFTNNQPPKVAHQGFQKIVNMKSENLFESQGGPI 173
Db   : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
445 ASASQACSFSDGYP-----LPSWTWKKCD-----KSPNCTE----- 476
QY 174 IMAIENEGPVWEIGCAPKAYTKWAAQMAVGLKTGVPWIMCKQEDAPDPVIDTCN--- 230
Db   ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
477 -----ETEGVWNRKANRKFQGVSSSTLNMSAIGFLVK-----CCAYNS 519
QY 231 -GFYCEGFRPNPKPKMTEVWTGMYTKFGGPIQORPAEDIAFSVARFVQNNGSFF--- 286
Db   ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
520 LGTSCETILLNSP-----GFPF-----FIQDNISFYATI 548
QY 287 -----NYMYHGGTNE-----GRTSSGLFIA--TSYDYDAPLD----- 317
Db   : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
549 GVCLLFIVVLTLLICHKKQFYESQLQMVQVTGSSDNEYFYVDFREYEDLKWEFPRE 608
QY 318 --EYG-----LLNEPKYGHLR---DLHKAIXL-----SEPALVSSYAAVTSL 354
Db   ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
609 NLEFGKVLGSGAFGKVMNATAYGISKTGYSIQVAVKMLKEKADSSEREALMSELMKMTQL 668
QY 355 GSNQ-----BAHYRSKSGACAFLSNY-----DSRYSVKVT 386
Db   ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
669 GSHENIVNLLGACTLSGPIYLIFEYCCYGDLLNLRKREKPHRTWTEIFKEHNFSEFYPT 728
QY 387 FQNRPNLPWISILPDCKTAVYNTAQVNSQSSSTKMTMPAGGGLSWQSYNEETPTADDS 446
Db   ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
729 FQSHPNSSMPGSRREV-----QIHPDSQI-----SGLHGNFHSF-----DEI 766
QY 447 DTLTANGLWEQKNV-TRDSSDYLWYMTNVTNVIASNEGFLKNGKDPYLTVMKAGHVLHVFVN 505
Db   : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
767 EYENQRRLREEDDLNVLTTEDLLCFA--YQVAKGMFELE--FKSCVHRDLAARNVL--VTH 821
QY 506 GKL-----SGTVYGTLDNPKLATYGNVKKLRAGINKISLLSVSGVGLPNVGVHYDTWAGVL 560
Db   ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
822 GKVKKICDFGLARDINSDSNVYVRGNARL--PVKWWAPESLFEGL--YTIKSDVWSYGIL 877
```

Search completed: June 24, 2002, 20:37:36
Job time: 222 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2002, 20:35:45 ; Search time 35.44 Seconds
(without alignments)
2269.115 Million cell updates/sec

Title: US-09-701-868-11
Perfect score: 3924
Sequence: 1 MLRTNVLLLVICLLDFSS.....VPEWGGNPTGISLRRSR 724

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues
Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0
Maximum DB seq length: 20000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : **US-09-701-868-11**
1: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3924	100.0	724	21	AA1980
2	2941	74.9	838	21	AA1981
3	2922.5	74.5	835	21	AA1982
4	2920	74.4	838	16	AA1983
5	2896	73.8	730	16	AA1984
6	2808.5	71.6	724	21	AA1985
7	2645.5	67.4	666	21	AA1986
8	2608	66.5	853	21	AA1987
9	2608	66.5	856	21	AA1988
10	2572	65.5	847	21	AA1989
11	2572	65.5	850	21	AA1990

12	2572	65.5	850	21	AA1991
13	2572	65.5	853	21	AA1992
14	2482.5	63.3	794	21	AA1993
15	2446.5	62.3	788	21	AA1994
16	2446.5	62.3	791	21	AA1995
17	2240.5	57.1	585	17	AA1996
18	2183.5	55.6	857	17	AA1997
19	2077	52.9	870	21	AA1998
20	1918	48.9	887	21	AA1999
21	978	24.9	251	21	AA2000
22	970	24.7	279	21	AA2001
23	951	24.2	274	21	AA2002
24	927	23.6	249	21	AA2003
25	738	18.8	295	21	AA2004
26	622	15.9	252	21	AA2005
27	536	13.7	636	22	AA2006
28	535	13.6	636	21	AA2007
29	531	13.5	636	20	AA2008
30	531	13.5	636	22	AA2009
31	531	13.5	636	22	AA2010
32	527	13.4	636	22	AA2011
33	509.5	13.0	637	21	AA2012
34	496	12.6	696	21	AA2013
35	495	12.6	677	19	AA2014
36	486	12.4	215	21	AA2015
37	485	12.4	668	22	AA2016
38	484.5	12.3	552	22	AA2017
39	483.5	12.3	654	22	AA2018
40	483.5	12.3	654	22	AA2019
41	483.5	12.3	654	22	AA2020
42	473	12.1	598	20	AA2021
43	470	12.0	123	21	AA2022
44	467.5	11.9	598	20	AA2023
45	461	11.7	687	22	AA2024

ALIGNMENTS

RESULT 1
ID AA1980 standard; protein; 724 AA.
AC AA1980;
XX 29-FEB-2000 (first entry)
XX Tomato beta galactosidase-4.
DE Tomato beta galactosidase-4.
KW fruit softening; beta galactosidase II protein; pectin;
KW transgenic plant; protoplast isolation.
XX Lycopodium esculentum.
XX Key Location/Qualifiers
FH Modified-site 282
FT Modified-site /note= "Possible N- glycosylation site"
FT Modified-site 459
FT Modified-site /note= "Possible N-glycosylation site"
FT Modified-site 713
FT Modified-site /note= "Possible N- glycosylation site"
FT Cleavage-site 23..24
FT /label= Leader_sequence_cleavage_site
XX WO9964564-Al.
XX 16-DEC-1999.
XX 08-JUN-1999; 99WO-US12697.
XX 09-JUN-1998; 98US-0088805.
XX

Arabidopsis thalia	850	21	AA1980
Arabidopsis thalia	853	21	AA1981
Arabidopsis thalia	794	21	AA1982
Arabidopsis thalia	788	21	AA1983
Arabidopsis thalia	791	21	AA1984
Arabidopsis thalia	585	17	AA1985
Arabidopsis thalia	857	17	AA1986
Nasturtium xylolu	870	21	AA1987
Tomato beta galact	887	21	AA1988
Tomato beta galact	251	21	AA1989
Tomato beta galact	279	21	AA1990
Arabidopsis thalia	274	21	AA1991
Arabidopsis thalia	249	21	AA1992
Arabidopsis thalia	295	21	AA1993
Arabidopsis thalia	252	21	AA1994
Human polypeptide,	636	22	AA1995
Hydrophobic domain	636	21	AA1996
Amino acid sequenc	636	20	AA1997
Human PRO polypept	636	22	AA1998
Human PRO236 prote	636	22	AA1999
Human polypeptide,	636	22	AA2000
Human Hydrolase pr	637	21	AA2001
Human prostate can	696	21	AA2002
Human beta galacto	677	19	AA2003
Arabidopsis thalia	215	21	AA2004
Portuguese Water d	668	22	AA2005
Human polypeptide,	552	22	AA2006
Human PRO polypept	654	22	AA2007
Human PRO262 prote	654	22	AA2008
Lactobacillus case	598	20	AA2009
Arabidopsis thalia	123	21	AA2010
Lactobacillus case	598	20	AA2011
Drosophila melanog	687	22	AA2012

(USDA) US DEPT OF AGRICULTURE.

Gross KC, Smith DL:

WPT: 2000-097532/08

N-PSDB: AAZ29341.

New beta-galactosidases, used to prepare transgenic plants with altered fruit ripening -

Claim 1; Fig 2; 85pp; English.

The present sequence is tomato beta galactosidase-4 (TBG-4) encoded by a cDNA derived from breaker, turning and pink fruit pericarp from 'Rugers' tomato plants. This hydrolyses terminal non-reducing beta-D-galactosyl residues from beta-D-galactosides leading to loss of tissue integrity and fruit softening. This is used for modifying cell wall metabolism and controlling ripening of fruit by altering activity of beta galactosidase II protein. Pectin with reduced galactosyl content is produced for use in biofilms or solutions. Transgenic plants with altered fruit ripening are produced by introducing DNA constructs comprising TBG cDNA. TBG forms a component of an enzyme mixture used to isolate protoplasts.

Sequence 724 AA:

Query Match	100.0%;	Score 3924;	DB 21;	Length 724;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 724;	Conservative	0;	Mismatches 0;	Indels 0;
Gaps	0;			

OV 1 MLRTNVLLLVICLLDEFFSSVKASVSYDDRAIIINGKRKILISGSIHYPRSTPOMWPDLI 60

1 mlrnlvlllviclllffssvksasvddralingkrlllsgslhvørstomwodd 60

OV 61 OKAKDGGI.DVJETYVFWNGHEPSPGKYNEEGRYDI.VRETKMVORAGI.YVNI.RIGPYVCAE 120

61 atakdag | dvjet vvf vnf g h e n s a k v n f e a r v d | v r f k m v a r a q | v v n j a n v v c a e | 120

0v 121 WNEGFPVWIKYVPGMEFFRTNNOPEKVA MGEFVKTVNMMKSENT.EESOCGPITMAOTEN 180

[illegible]

Qv 181 EVGPVFEI GAPGKAYTKWAAQMAVGLKTCVPWIMCKQEDAPDPVIDTCTNCFYCEGERBN 240

[illegible]

0v 241 KPYKPKMNTFVWTCGWYTKFCCGRTPOBPBAEDTAFSIVABEVONNCSEFENVVYUHCCTNECPT 300

Q7	241	KFIFNFWWLEVWIOWLINAGOFIFGFAPEDJAF SVARE VQNNQSSEFN IMLNGGINF GR1	300
D6	241	KGVKPKMFEWTGTATWKFGAIPARPAEDIAFSVARFYANNAGSFNVVMVHAATNFART	300

0v 301 SSCLEIATSVNDVDPRI DEVC I NERKVCHI BDI HKATKI CERAI VSEVALITSI CENOEFA 360

VY	307	SSGUE IATISUDAPUWUGLLENFRGHKJULNNAIALSEFADVSSIAVAVLSSQNGEA
Dh	301	ssmfiatsvvdanidevmllnenkvghlrdlhkaktlsenavsvsvaaaglnennea

361 HUYBEKSCALLEY ENYDERVSWITTEONBRUNY DREISTETI DDCVTAUWNTAUWNSCEE 130

301 HVTKRSKSGACAAFLSNSTDSKISVAVFQKRFNPLPPWSLSLLPDKTAVINTIAQVNSQSS 420
QY
361 huursksgaaaf lsnstdrvrukvut fgnrtvaal nruv is l l n d k t a v i n t i a q v n s q s s 420
Db

03 421 STAMPA CCI SNO SVNEE E E N D D C D E T E A N C I M E O V A N T E D C S D V I E V M E N T I T A C N E 180

[illegible]

401 CBT MAY BE MORE EFFECTIVE FOR ANXIETY AND DEPRESSION IN THE
CITY OF MICHIGAN

481 GFLEKNGRDPYLTVMSAGHVLHFVNGKLSGRVYGTLDNPRLKTYSGNWKLRAGINKISLES 540

1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525

QY 541 VSVGLPNVGVDYDTNAGVLGPTVLSGLNEGSRNLAKQWSYKVKVLKGESLSLHSLGSS 600

601 SVENVGSLMAQKQPLTWYKATFNAPGGNDPLALDMSMGKGQIWNINGEVRHWPGYIA 660

```
Db 6 ilmlnvlvl---lgswwfsqtasvdydhralivngqrllisgsvhyprstpmwpgii 62
Qy 61 QKAKDGLDLVETVYFWNGHEPSPGKYNFEGRYDLVRFKIMVORAGLVNLRIGPYCAE 120
Db 63 qkakegvdvltqyvfwnghhepgqgkyfyegrydlvkfkivhdaqlyvhlrvpyacae 122
Qy 121 WNGGFPVWLKYVPGMEFRTNNQPKFVAMQGFVQKIVNMKSENIFESOGGPIIMAIEN 180
Db 123 wnggfpvwlkyvpgisfrtdngpfkaamqkftakivnmkaerlyetqggpilisqien 182
Qy 181 EYGPVWEIGAPGKATYTKWAQMAVGLKTGYPWIMCKQEDAPDPVIDTCNGFYCEGFRPN 240
Db 183 eygmewelgapgksyaqaamavldtgvpmwckqddapdplinaacngfyedyspn 242
Qy 241 KPYKPMTEVMTGWYTKFGGPIPORPAEDIAFSAFVARVQNGSFENYMYHGGTFNGRT 300
Db 243 kaypkivteawtftqgnpvyrpaedlafsvakfiqkggsfinymyhggtntfirt 302
Qy 301 SSGLFIAFSDYDAPLDYGLLNEPKYGHRLDLHKAIKLSEPALVSSYAAVTSLSGNSQEA 360
Db 303 aggpfiatsydydapldeyglirgpkwghldlhraiklcepalvsgdpavtaighqgea 362
Qy 361 HVYRSKSCACAFLSNYSRYSVKTFONRPNPLPPWISILPDCKTAVYNTAQVNSOSS 420
Db 363 hvfrskagscaafianyqdhfsfatvstfanrhynlppwsisilpdckntvntarigaqsa 422
Qy 421 SIKMTPAGGGLSWOSYNBETPTADDSDTLTANGLEWQKNVTRDSSDYLWYMTNVNVIASNE 480
Db 423 qmktptvsgripwgsfneetsyeds-sftvvgllqeqinttrdvsdylwystdckidre 481
Qy 481 GFLKNGKDPYLTVMASGHVLHVFNVGKLSGTVYGTFLDNPKLTYSGNVKLKLAGINKISLLS 540
Db 482 kflrggkwpwltmsaghalhvfngqgiagtaygslekpkltfsvkavnlragvknkisl 541
Qy 541 VSVGLPNVGVHYDTWNAAGLVPTVLSGLNEGRNLAKQWYKVLKCESLSLSLSSGSS 600
Db 542 iavglpnlgpfetwneagvlpgvsltdigedkrdltwqkwykvgkgealslsdsgss 601
Qy 601 SVEWVRSLMAQKOPLTWYKATFNAPGNDPLALDMSMGQGIWINGEGVGRHWPGYIA 660
Db 602 sveewegslvaqrpltwyktfnapagnndplaldnltmgkgqwiwngsglrywpgyka 661
Qy 661 QGDCSKSYAGTFNEKKCQTNCGQPSQRWYHVPRSWLKPNSNLLVVFSEWGNPFGISLV 720
Db 662 sgncgacnyagwfnckclscgeasgrwyhvprswlyptgnllvlfeewggegpgislv 721
Qy 721 RR 722
Db 722 kr 723

RESULT 3
AAY44303
ID AAY44303 standard; protein; 835 AA.
XX -
AC AAY44303;
XX -
DT 29-FEB-2000 (first entry)
XX -
DE Tomato beta galactosidase-1.
XX
KW Tomato beta galactosidase-1; TBG; Rutgers tomato plant; pectin;
KW fruit softening; beta galactosidase II protein; biofilm;
KW transgenic plant; protoplast isolation.
XX
OS Lycopersicon esculentum.
XX
FH Key Location/Qualifiers
FT Peptide 1..24
FT /label= Signal_peptide
FT Protein 25..835
FT /label= beta-galactosidase-1
XX
```

```
PN W09964564-A1.
XX
PD 16-DEC-1999.
XX
PF 08-JUN-1999; 99WO-US12697.
XX
PR 09-JUN-1998; 98US-0088805.
XX
PA (USDA ) US DEPT OF AGRICULTURE.
XX
PI Gross KC, Smith DL;
XX
DR WPI; 2000-097532/08.
DR N-PSDB; AAZ29338.
XX
PT New beta-galactosidases, used to prepare transgenic plants with altered
XX fruit ripening -
FS Claim 1; Fig 2; 85pp; English.
XX
CC The present sequence is tomato beta galactosidase-1 (TBG-1) encoded by a
CC cDNA derived from breaker, turning and pink fruit pericarp from 'Rutgers'
CC tomato plants. This hydrolyses terminal non-reducing beta-D-galactosyl
CC residues from beta-D-galactosides leading to loss of tissue integrity and
CC fruit softening. This is used for modifying cell wall metabolism and
CC controlling ripening of fruit by altering activity of beta galactosidase
CC II protein. Pectin with reduced galactosyl content is produced for use in
CC biofilms or solutions. Transgenic plants with altered fruit ripening are
CC produced by introducing DNA constructs comprising TBG cDNA. TBG forms a
CC component of an enzyme mixture used to isolate protoplasts.
XX
SQ Sequence 835 AA;
Query Match 74.5%; Score 2922.5; DB 21; Length 835;
Best Local Similarity 72.2%; Pred. No. 1.le-239;
Matches 518; Conservative 94; Mismatches 102; Indels 3; Gaps 2;
Qy 6 VLLLVLCILDPFSSVKASVSVDRAITINGKRKILISGSIHVPRSTPQMWPDLIQKAD 65
Db 7 mlmlmlcl--wvscgiasvdydhraivngqrllisgsvhyprstpmwpdliqkake 64
Qy 66 GGLDVIETVYFWNGHEPSPGKYNFEGRYDLVRFKIMVORAGLVNLRIGPYCAEAFEGG 125
Db 65 ggvdvltqyvfwnghhepgqgkyfyegrydlvkfkivhdaqlyvhlrvpyacaewnfgg 124
Qy 126 FPVWLKYVPGMEFRTNNQPKFVAMQGFVQKIVNMKSENIFESOGGPIIMAIENEYGPV 185
Db 125 fpvwlkyvpgisfrtdngpfkaamqkftakivnmkaeklyetqggpilisqieneygpm 184
Qy 186 EWEIGAPGKATYTKWAQMAVGLKTGYPWIMCKQEDAPDPVIDTCNGFYCEGFRPNKPKY 245
Db 185 eweigepkkyvsewaakmavldtgvpmwckqddapdplintcngfyedytfnkankp 244
Qy 246 KMTEVMTGWYTKFGGPIPORPAEDIAFSAFVARVQNGSFENYMYHGGTFNGRTSSGGLF 305
Db 245 kmwteawtawtftqgnpvyrpaedmafavarfiqtggsfinymyhggtntfirtsggpf 304
Qy 306 TATSVDYDAPLDYGLLNEPKYGHRLDLHKAIKLSEPALVSSYAAVTSLSGNSQEAHVRS 365
Db 305 tatsydydapldeyglirgpkwghldlhraiklcepalvsvdptvtslgnqgearvfks 364
Qy 366 KSGACAFLSNYSRYSVKTFONRPNPLPPWISILPDCKTAVYNTAQVNSOSSSIKMT 425
Db 365 esgacaaflanyqhsvfakvafgmhynlppwsisilpdckntvntarvsgsqmkm 424
Qy 426 PAGGLSQSWEETPTADDSDTLTANGLEWQKNVTRDSSDYLWYMTNVNVIASNEGFLKN 485
Db 425 pvsrgfswesfneedaashed-dftvvgllqeqinttrdvsdylwystdiedtpeglins 483
Qy 486 GKDPYLTVMASGHVLHVFNVGKLSGTVYGTFLDNPKLTYSGNVKLKLAGINKISLSVSVGL 545
Db 484 gnwpwltvtsaghalhvfngqgiagtaygslenpkltfsgnlnragvknkislslavgl 543
```


PI Chengappa S, De SILVA J, Hellyer SA, Reid JSG;

XX WPI; 1995-311537/40.

DR N-PSDB; AAT01014.

XX

PT Lupin and tomato exo-galactanase DNA and protein - useful for the
PT alteration of plant characteristic(s), e.g. texture, growth,
PT ripening

XX

PS Claim 1; Page 25; 67pp; English.

XX

CC This enzyme may be expressed in a transgenic plant in order to
CC alter characteristics of e.g. growth, texture or ripening of the
CC plant or plant parts. This protein may also be expressed in a
CC recombinant host for the production of the enzyme which may be used
CC for the modification, degradation or liquefaction of plant materials
CC in order to affect mechanical properties relating to eating texture,
CC particle sizes of, e.g. fruit or vegetable juices, or extractability
CC of colors, flavors or vitamins.

XX

SQ Sequence 730 AA;

Query Match 73.88; Score 2896; DB 16; Length 730;

Best Local Similarity 71.4%; Pred. No. 1.6e-237;

Matches 520; Conservative 94; Mismatches 100; Indels 14; Gaps 5;

QY 1 MLRTN---VLLLV--ICLLDFSSKASVSDDRAIINGKRIKILISGSIHYPRSTPQM 55
DB 12 msrrnfmvllllffwcy-----vtasvtydhkamingqrilllsgsihyprstpqm 65
QY 56 WPDLIQAKDGGLDVIETTYFWNGHEPSPGKYNFEGRYDLVRFIKMVQAGLYNLRIGP 115
DB 66 wpdliqakdgglvdietyfwnghepspgkyfedrldlvgiklvqaglfvhlrigp 125
QY 116 YVCAEWNFGFPVWLKVPVGMFEFTNQPKVAMQGGFVQKIVNMKSENLFESOGGPIIM 175
DB 126 ficawefgfpvwlkvpvgvgaftndneptkeamqkitekivlmkaeklfqsgggplil 185
QY 176 AOINETGPPVEWEIGAPKATYKAAQMAVGLTKVPMCKQEDAPDPVDTGNGFYCE 235
DB 186 sqleneyppveweigapkgatkykwaagmavldtgvpmckqgedaldpdiidtcngfyce 245
QY 236 GFRENKPKYKPMWTEVMTGWTKYFEGGPIPORPAEDIAFSAVFQNNGSFENYMYRGGT 295
DB 246 nftpnknykpkltewtgytafggatpyrpaediafsvafiqnrgslfnymyhggt 305
QY 296 NFGTSSGLFIATSYDYDAPLDEYGLNEPKYGHRLDRAIKLSEPALVSSYAATVSLG 355
DB 306 nfgtssnglfiatasydydapideyglnepkghrlraiklseqsalvsvdptvswpg 365
QY 356 SNOEAHYVRKSGACAFLSNYDSRYSVKVTFOKRPVNLPSWISILPDCKTAVYNTAOV 415
DB 366 knlevhlyktes-acaaflanynctdytqvkfngqdyldpwsisilpdcktevfntakv 424
QY 416 NSQSSSIKMTAGGLSQWSEETPTADDSDTLTANGLEQKVNTRDSSDYLWYMTNVN 475
DB 425 nsplrhrkmtptvnsafawqsyneepssendpvtgylwqevgvttrdsdylyltdvn 484
QY 476 IASNEGFLKNGKDPYLIVMSAGHVLHFVNGKLSGTGYGLDNLKPLTYSGNVKLKAGINK 535
DB 485 igpnd--ikdgkwpvlamsaghlvningqyagtagysiddprltfsgqvnlrvggnk 542
QY 536 ISLLSVSGLPVGHVHVDTNAGVLPVTLTSLNGESRNLAKORWSKVGLKGESLSLHS 595
DB 543 isllsvsvglanvgchretwnvgpvtlctgssgtdwtskqkwsykgikgesislt 602
QY 596 LSGSSSVWVRGSLMAOKQPLTYKATFNAPGNDPLALDMSMGKQIWTNGEGVGRHW 655
DB 603 eagsnsvwvrgslvakkqlawktlfsapagnldlsgngkgvwnvgqsigrhw 662
QY 656 PGYTAQGDCKSCSYAGTFNEKKCTQNCQPSQRVHYHVPKPSGNLLVVFEEWGNPT 715
DB 715 pgytaqgdckscsyagtfnekkctqncqpsqrhyhvpkpsgnllvvfveewgnpt 715

Db 663 pgnkargncnnyagtytdtkclancgqpsqrwyhprwlrsggnylvvleewgdpn 722
QY 716 GISLVRRS 723
DB 723 gialvert 730
RESULT 6
AAG49309
ID AAG49309 standard; Protein; 724 AA.
XX AAG49309;
AC AAG49309;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 62368.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.

PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.

PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 71.68; Score 2808.5; DB 21; Length 724;

Best Local Similarity 70.2%; Pred. No. 4.4e-230;

Matches 503; Conservative 92; Mismatches 119; Indels 3; Gaps 3;

Qy 7 LLLVLCILDFSSVKASVSDRAIINGKRKILISGSIHYPRSTPOMWDLIOKAKDG 66

Db 11 iflailccslscivkasvsydrkavlingrrllsgsihyprstpmwpgliqakeg 70

Qy 67 GLDVIETYVFWNGHEPSPGKYNFEGRYDLVRFIKWQVRAGLYVNLRIIGYVCAEWNFGGF 126

Db 71 gldvetyvfwnghepspggyfdrydlvklivhdaglyvnlrigrpyycaewnfggf 130
Qy 127 pwlklyvpgmefrnnopfkvamogfvokivnmkksenlfesogqpiimaotieneygpe 186
Db 131 pwlklyvpgmafrdnepfkamkftekiwmmkaeklfqgqgpiillaenierygpe 190
Qy 187 weigapgaytkwaamaavglktgvpwimckoeapdvipdctngfycgegrprnkpypk 246
Db 191 weigapgaytkwaamaavglktgvpwimckoeapdvipdctngfycgegrprnkpypk 250
Qy 247 mtevwgwytkfkgpripdorpaeadiarsvarfvqngsffnyvmyhgtngftrssglfi 306
Db 251 mwtenwtgwytdfggavpyrpediayvarfiqggsilvnyymyhggtntfdr- agefm 309
Qy 307 atsydydapldexyglpnykylhldhkaiklsepavlvssvavtsglsnoeahvyrsk 366
Db 310 assydydapldexyglpnykylhldhkaiklsepavlvssvavtsglsnoeahvyrsk 369
Qy 367 sgacaafslsnydsrysvkvtpqnrpylnppwsiisilpdkrtavvntaqvnsqssikmtp 426
Db 370 s-scaafslsnydsrysvkvtpqnrpylnppwsiisilpdkrtavvntaqvnsqssikmtp 428
Qy 427 agglswosyneerptaddstflranglweoknvtrodsdylwtvntvniassnegflkng 486
Db 429 tgtkfwsgsfneatpneagtfnarglveqtsmtwkdskdyfwitdtigsetflktg 488
Qy 487 kdpvltvmsaghlvfvngklsgvtygtldnplklytsnkvklraginkisllsvsglpl 546
Db 489 dsplltvmsaghlvfvngklsgvtygtldnplklytsnkvklraginkisllsvsglpl 548
Qy 547 nvghydtvwnagvlpgvptlsglnegrnlakokvskvylkgeslshslsgssvewvr 606
Db 549 nvghydtvwnagvlpgvptlsglnegrnlakokvskvylkgeslshslsgssvewvr 608
Qy 607 gslmaqopltykatfnagpndplaldmasmgkqgiwingevgvgrhwpqyiaogdcsk 666
Db 609 gsfvakkpqltykstfatpagnelaldmtmgkgvwiingrigrhwpaykagqscgr 668
Qy 667 cysagtfnkctocnccgopsorvhpvrswlkspsnllvlfvfeemgngptglsilvrrs 723
Db 669 cnyagtfdakkclscnccgopsorvhpvrswlkspsnllvlfvfeemgngptglsilvrrs 724

RESULT 7
AAG49310
ID AAG49310 standard; Protein; 666 AA.
XX AC AAG49310;
XX DT 18-OCT-2000 (first entry)
XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 62369.
XX KW Protein identification; signal transduction pathway; metabolic pathway;
XX KW hybridisation assay; genetic mapping; gene expression control; promoter;
XX KW termination sequence.
XX OS Arabidopsis thaliana.
XX PN EP1033405-A2.
XX PD 06-SEP-2000.
XX PF 25-FEB-2000; 2000EP-0301439.
XX PR 25-FEB-1999; 99US-0121825.
XX PR 05-MAR-1999; 99US-0123180.
XX PR 09-MAR-1999; 99US-0123548.
XX PR 23-MAR-1999; 99US-0125788.
XX PR 25-MAR-1999; 99US-0126284.
XX PR 29-MAR-1999; 99US-0126785.
XX PR 01-APR-1999; 99US-0127462.

PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.

PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145919.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 04-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 04-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.

PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 67.4%; Score 2645.5; DB 21; Length 666;

Best Local Similarity 70.7%; Pred. NO. 2.9e-216;

Matches 473; Conservative 83; Mismatches 110; Indels 3; Gaps 3;

Qy 55 MWPLIQAKDGLDIETVYFVWNGHSPGKYNEFGRYDLVRFIKMVRAGLYVNLRTG 114
Db 1 mwpqliqakdegldvetyfvwnghepspgdyfyfgydyldvfkfklvhaqlyvnlri 60
Qy 115 PYVCAEWNFGFPVWLKYVPGMEFTNNQPFKVMQGVQKIVNMNMSENLEFESQGPTI 174
Db 61 pyvcaewnfgfpvwlkfvpgmaftrdnepfkaamkkftekivmmkaeklfqtaggpli 120
Qy 175 MAQIENEYGPVEWEIGAPGKAYTKWAAQMAVGLKTGVPMWCKQEDADPPVIDTCNGFYC 234
Db 121 laqleneygpvewei gapgkaytkwaaqmaiglstgvpmckqedapppiidtcngyyc 180
Qy 235 EGFRENKPKYKPMWTEVWTGYTKFGGPIQORPAPEDIAFVSARFVQNNQSFNYYMYHGG 294
Db 181 edfkpsinkpkmwtenwtgytdfggavpyrpvediasyvarfiqkggslvnyymyhg 240
Qy 295 TNFGRTSSGLFIATSYDYADPLDEYGLLNEPKYGHRLDHLKAIKLSEPALVSSYAAVTSL 354
Db 241 tnfdrt-afefmassydydapldeyglprepkyslhkaikalsepalisadatvtsl 299
Qy 355 GSNQEAHVYRSKGACAAFLSNYDSRYSVKYTFQNPYNLPWISILPDCKTAYVNTAQ 414
Db 300 gakeayvfwsk-scaaflnknkdenaarvlfirgfydlppwsvsilpdcckteyvtak 358
Qy 415 VNSQSSSIKMTAGGSLWSQSYNEETPTADSDTITANGLWEQKNVTRDSSDYLMYMTNV 474
Db 359 vnapsvhrnmvptgkfkfsgfneaptaneagtfainglveqismtdwksdyfywidi 418
Qy 475 NIASNEGFLKNGKDPYLTVMASAGHVLHVFNKGLSGTYVYGTLDNPKLYSGNVKLRAGIN 534
Db 419 tigsetflktgdsplltvmaaghalhvfngqlsgtayggldhpkltfsqkikhagv 478
Qy 535 KISLLSVSGLPNVGVHYDTWAGVGLPVYTLISGLNEGSRLAKQKWSYKVLKGLKGSLSLH 594
Db 479 kiallsvavglpnvgtfheqwnkgvlgpvtlkgvnsfgtdmskwwksykgvkgkgealslh 538
Qy 595 SILSGSSSVIEWVRGSLMAOKPLTWYKATFNAPGGNDPLALDMSMGKQGIWINGGVGRH 654
Db 539 tntessgvrtqgsvfakkqpltwykstfapagnepaladmntmgkgqvwingnigrh 598
Qy 655 WPGYIAQGDCKSCSYAGTFNEKKQTCQNGQPSQRWYHVPRLKPSGNLLVVVFEWGNP 714

```

|| | ||| | :|:||||: ||| :|||: ||||| ||||| | ||:|||| ||:|
Db 599 wpaqagsgcgrcnaygtfdakkclsnceasgrwyhvrswlk-sgnliivfeelggdp 657
QY 715 TGISLVRRS 723
    |||||:|
Db 658 ngislvkrt 666

RESULT 8
AAG49423
ID AAG49423 standard; Protein; 853 AA.
XX
AC AAG49423;
XX
DT 18-OCT-2000 (first entry)
DE
DE Arabidopsis thaliana protein fragment SEQ ID NO: 62523.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD
PD 06-SEP-2000.
XX
XX 25-FEB-2000; 2000EP-0301439.
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
```

PR	12-AUG-1999;	99US-0148341.	
PR	13-AUG-1999;	99US-0148565.	
PR	13-AUG-1999;	99US-0148684.	
PR	16-AUG-1999;	99US-0149368.	
PR	17-AUG-1999;	99US-0149175.	
PR	18-AUG-1999;	99US-0149426.	
PR	20-AUG-1999;	99US-0149722.	
PR	20-AUG-1999;	99US-0149723.	
PR	20-AUG-1999;	99US-0149929.	
PR	23-AUG-1999;	99US-0149902.	
PR	23-AUG-1999;	99US-0149930.	
PR	25-AUG-1999;	99US-0150566.	
PR	26-AUG-1999;	99US-0150884.	
PR	27-AUG-1999;	99US-0151066.	
PR	27-AUG-1999;	99US-0151080.	
PR	30-AUG-1999;	99US-0151303.	
PR	31-AUG-1999;	99US-0151438.	
PR	01-SEP-1999;	99US-0151930.	
PR	07-SEP-1999;	99US-0152363.	
PR	10-SEP-1999;	99US-0153070.	
PR	13-SEP-1999;	99US-0153758.	
PR	15-SEP-1999;	99US-0154018.	
PR	16-SEP-1999;	99US-0154039.	
PR	20-SEP-1999;	99US-0154779.	
PR	22-SEP-1999;	99US-0155139.	
PR	23-SEP-1999;	99US-0155486.	
PR	24-SEP-1999;	99US-0155659.	
PR	28-SEP-1999;	99US-0156458.	
PR	29-SEP-1999;	99US-0156596.	
PR	04-OCT-1999;	99US-0157117.	
PR	05-OCT-1999;	99US-0157753.	
PR	06-OCT-1999;	99US-0157865.	
PR	07-OCT-1999;	99US-0158029.	
PR	08-OCT-1999;	99US-0158232.	
PR	12-OCT-1999;	99US-0158369.	
PR	13-OCT-1999;	99US-0159293.	
PR	13-OCT-1999;	99US-0159294.	
PR	13-OCT-1999;	99US-0159295.	
PR	14-OCT-1999;	99US-0159329.	
PR	14-OCT-1999;	99US-0159330.	
PR	14-OCT-1999;	99US-0159331.	
PR	14-OCT-1999;	99US-0159637.	
PR	14-OCT-1999;	99US-0159638.	
PR	18-OCT-1999;	99US-0159584.	
PR	21-OCT-1999;	99US-0160741.	
PR	21-OCT-1999;	99US-0160767.	
PR	21-OCT-1999;	99US-0160768.	
PR	21-OCT-1999;	99US-0160770.	
PR	21-OCT-1999;	99US-0160814.	
PR	21-OCT-1999;	99US-0160815.	
PR	22-OCT-1999;	99US-0160980.	
PR	22-OCT-1999;	99US-0160981.	
PR	22-OCT-1999;	99US-0160989.	
PR	25-OCT-1999;	99US-0161404.	
PR	25-OCT-1999;	99US-0161405.	
PR	25-OCT-1999;	99US-0161406.	
PR	26-OCT-1999;	99US-0161359.	
PR	26-OCT-1999;	99US-0161360.	
PR	26-OCT-1999;	99US-0161361.	
PR	28-OCT-1999;	99US-0161920.	
PR	28-OCT-1999;	99US-0161992.	
PR	28-OCT-1999;	99US-0161993.	
PR	29-OCT-1999;	99US-0162142.	
Query Match			66.5%; Score 2608; DB 21; Length 853;
Best Local Similarity			63.1%; Pred. No. 6.9e-213;
Matches 457; Conservative 119; Mismatches 136; Indels 12; Gaps 5;			
QY	8	LLLVICL-----LDFSSVKASVYDDRAIIINGKRKILISGSIHYPRSTPQWMPDLIQ	61
Db	10	lilwfcglflilgvvf---vqcgvydrkallingqrllfsgsihyprstpdmdwediql	66

QY	62	KAKDGLDVIETYYVFWNGHEPSPGKYNFGRYDLVRFIMQVORAGLYVNLRTGPPYCAEW	121
Db	67	kakdgidvietyvfwnlhhpspgkydfegrrndrfvktihkaglyahrlppyycaew	126
QY	122	NFGGPPVWLKYYPGMEFTNNQPEKVMQGFVOKTVNMKSENLPESOGPIIQAQIENE	181
Db	127	nfggfpvwlkyvpgisfrtdnepfkramkgfterivelmksenlfesggppllsqiene	186
QY	182	YGPVEWEIGAPGKAYTKWAAOAVGLKTGVPWIMCKQEDADPVPIDTCNGFYCEGFRPNK	241
Db	187	ygrqgllgaeghnymtwaakmaiatetgypvwmcckeddpdvintcngfycdsfapnk	246
QY	242	PYKPKMTEVWTGWYTKFGGPIQORPAEDIAFSVARFVQNNGSFFNYMYHGGTNGFRTS	301
Db	247	pykpllwteawsgwtfegfpmhhprvqdlafgvarfigkggsvfnyymyhgtnfgrta	306
QY	302	SGLFIATSYDYDAPLDEYGLLNEPKYGHLDLHKAIKLSEPALVSSYAAVTSILGSNOEAH	361
Db	307	ggpfvttsydydapideyglirgpkylhkeihraikmcekalvsadpvtvtsignkqgah	366
QY	362	VYRSKSGCAAFLSNVDSRYSVKVTQFONRPNYLPWMSISILPDCKTAVYNTAQNVSQSS	421
Db	367	vysaesgdcsaflydytesaarvifnnvhnlpwmsisilpdcnavntakvgvqtseq	426
QY	422	IKMTPAG-GGLSQSYNEETPTADSDTLTANGLEQKNVTRDSSDYLMYMTNVNIASNE	480
Db	427	memlptdtknfqwesyiledlsslddsstftbllgeqinvtrdtsdylwymtsvdigdse	486
QY	481	GFLKNGKDPYLTVMAGVHLHVFAVNGKLSGTVYGTLDNPKLTYSGNVKLRAGINKISLLS	540
Db	487	sflhggeiptlilqstghavhifvngqlsgsafrqrnrfrtyqgklnhsgtnralls	546
QY	541	VSVGLPNVGHYDWTNAGVLGPVTLISGLNEGRNLAKQKSYKVLKGBSLSLHLSGSS	600
Db	547	vavglpnvghfeswntgilgpvalhglsgqkmdlswkwtvyqvgkkgaeamlaftntp	606
QY	601	SVEVYRGSLSMAQK-QPLTWYKATFNAPGNDPLALDMSMGKGQIWIINGEGVRRHWPYI	659
Db	607	sigmdasiltvqkqpqltwktyfdapegneplaidmegmgkgqiwnngesigrywtaf-	665
QY	660	AOGDCSKSYAGTFNEKKCOTNCGOPSORWYHVPKSLKPSGNLAVFPEEMGNPTGISL	719
Db	666	atgdcshcsygtgtykpnkcqtcgqgptqrwyhvprawlpxsqnllvifeelggnpstvs	725
QY	720	VRRS 723	
Db	726	vtrs 729	
RESULT 9			
AAG49422			
ID	AAG49422 standard; Protein; 856 AA.		
XX	AAG49422;		
XX	18-OCT-2000 (first entry)		
DE	Arabidopsis thaliana protein fragment SEQ ID NO: 62522.		
XX	Protein identification; signal transduction pathway; metabolic pathway;		
KW	hybridisation assay; genetic mapping; gene expression control; promoter;		
KW	termination sequence.		
OS	Arabidopsis thaliana.		
XX	EPI033405-A2.		
PD	06-SEP-2000.		
XX	25-FEB-2000; 2000EP-0301439.		
PR	25-FEB-1999; 99US-0121825.		
PR	05-MAR-1999; 99US-0123180.		

PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 17-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.

PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 16-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.

PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148664.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 13-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 26-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match		65.5%; Score 2572; DB 21; Length 847;
Best Local Similarity		62.0%; Pred. No. 8e-210;
Matches 453; Conservative 118; Mismatches 128; Indels 32; Gaps 6;		
QY	8 LLLVLCI-----LDFFSVKASVSDRAIIINGKRKILISGSIHYPSTPQWMPDLIQ	61
Db	10 lllwfcglflllgvf---vqcgvydrkallingrrllfsgsihprstpdmdwllq	66
QY	62 KAKDGLDVIETVYPWNGHEPSPGKYNEGVRDLVRFKTMQVAGLVNLRIGPVVCAEW	121
Db	67 kdkdgiadietvynlhnepspgkydregndlvrfvktlhnkaglyanhrigpyvcaeW	126
QY	122 NFGGPPWLKYVPGMEFRTNQPKVAMQGVQKIVNMNMKSENLFESOGGPIIIMAAIQENE	181
Db	127 nfggppwlkyvpgisfrtdnepkramkgfterivelmkseinfesggpillsqiene	186
QY	182 YGPWEWEIGAPGKAYTKWAAQAVGLKTGVPMCKQBDADPPVIDTCNGFYCGFRPNK	241
Db	187 ygrgqlllgaeghnymtwaakmaiatetgvpwmckedadpvpintcngfydsfapnk	246
QY	242 PYKPMWTEVMTGWYTKFGGPIPORPAEDIAFVSARFVQNNGSFENYMYHGGTFNFGRTS	301
Db	247 pykplwteawsgwftfeggmhhprpvgdlafgvarfikggsfvnyymyhggtafgrta	306
QY	302 SGLFATSYDYDAPLDEYGLLNEPKYGHRLDLKAIKLSEPALVSSYAAVTSLSGNQB--	359
Db	307 ggpfttsydydapideyglirqpkyghikelhraikncekalysadpvtsgnkgqvw	366
QY	360 -----AHVYRSKSGACAAFLSNDSDRSYKVTQNRPNYLPWWSISILPCKTAVYNNTA	413
Db	367 lyyerfahvysaesgdcsafianydytesaarvlnnnvhnlpwvssisilpdcnnavfnta	426
QY	414 QVNSQSSSIKWTGAGGLSWQSYNEETPTADSDTLTANGLWEOKNVTRODSSDYLWYMTN	473
Db	427 kvsn-----fqwesyledssldsdstftchleglnvtrcdtsdylwymts	473
QY	474 VNIAINEGFLNKDPLYTVMSAGVHLHVFNGLSGFVYGTLDNPKILYSGNVKLRAGI	533
Db	474 vdigdsesflhggepltiilgstghavhifvngqlsgsafqrnfrytgkkinlhsqt	533
QY	534 NKISLLSVSGLPNVGHYDTWNAAGVLQPVYLSGLNEGSRNLAKQKSYKVLKGESLSL	593
Db	534 nriallsavaglpnvghfeseswntgilpvaalhgsqgkmdslwqkwtvgvlgkgeamnl	593
QY	594 HSLGSSSVWVRGSLMAOK-QPLTWYKATFNAPGNDPLALDMSMGKGQIWIINGEGVG	652
Db	594 afptnpsigmmdasltvqkppqtlwhtkyfdaepgneplaldmegmkqgiwvngesig	653
QY	653 RHPGPIYIAGDCSKCSYAGTNEKKCQTNCGOPSORWYHVPRLKPSGNLLVVFEEWGG	712
Db	654 rywtaf-atgdcshcsytgtykpnkcqtcgqgptqrwyhvprawlkpsqnlvifeelgg	712
QY	713 NPTGSLVRRS 723	
Db	713 npstsvlvkrs 723	
RESULT 11		
AAG31161		
Id	AAG31161 standard; Protein: 850 AA.	
XX		
AC	AAG31161;	
XX		
DT	17-OCT-2000 (first entry)	
XX		
DE	Arabidopsis thaliana protein fragment SEQ ID NO: 37376.	
XX		
KW	Protein identification; signal transduction pathway; metabolic pathway;	
KW	hybridisation assay; genetic mapping; gene expression control; promoter;	
KW	termination sequence.	
XX		
OS	Arabidopsis thaliana.	

XX	EP1033405-A2.
PN	
XX	
PD	06-SEP-2000.
XX	
PF	25-FEB-2000; 2000EP-0301439.
XX	
PR	25-FEB-1999; 99US-0121825.
PR	05-MAR-1999; 99US-0123180.
PR	09-MAR-1999; 99US-0123548.
PR	23-MAR-1999; 99US-0125788.
PR	25-MAR-1999; 99US-0126284.
PR	29-MAR-1999; 99US-0126785.
PR	01-APR-1999; 99US-0127462.
PR	06-APR-1999; 99US-0128234.
PR	08-APR-1999; 99US-0128714.
PR	16-APR-1999; 99US-0129845.
PR	19-APR-1999; 99US-0130077.
PR	21-APR-1999; 99US-0130449.
PR	23-APR-1999; 99US-0130510.
PR	23-APR-1999; 99US-0130891.
PR	28-APR-1999; 99US-0131449.
PR	30-APR-1999; 99US-0132048.
PR	30-APR-1999; 99US-0132407.
PR	04-MAY-1999; 99US-0132484.
PR	05-MAY-1999; 99US-0132485.
PR	06-MAY-1999; 99US-0132486.
PR	06-MAY-1999; 99US-0132487.
PR	07-MAY-1999; 99US-0132863.
PR	11-MAY-1999; 99US-0134256.
PR	14-MAY-1999; 99US-0134218.
PR	14-MAY-1999; 99US-0134219.
PR	14-MAY-1999; 99US-0134221.
PR	14-MAY-1999; 99US-0134370.
PR	18-MAY-1999; 99US-0134768.
PR	19-MAY-1999; 99US-0134941.
PR	20-MAY-1999; 99US-0135124.
PR	21-MAY-1999; 99US-0135353.
PR	24-MAY-1999; 99US-0135629.
PR	25-MAY-1999; 99US-0136021.
PR	27-MAY-1999; 99US-0136392.
PR	28-MAY-1999; 99US-0136782.
PR	01-JUN-1999; 99US-0137222.
PR	03-JUN-1999; 99US-0137528.
PR	04-JUN-1999; 99US-0137502.
PR	07-JUN-1999; 99US-0137724.
PR	08-JUN-1999; 99US-0138094.
PR	10-JUN-1999; 99US-0138540.
PR	10-JUN-1999; 99US-0138847.
PR	14-JUN-1999; 99US-0139119.
PR	16-JUN-1999; 99US-0139452.
PR	16-JUN-1999; 99US-0139453.
PR	17-JUN-1999; 99US-0139492.
PR	18-JUN-1999; 99US-0139454.
PR	18-JUN-1999; 99US-0139455.
PR	18-JUN-1999; 99US-0139456.
PR	18-JUN-1999; 99US-0139457.
PR	18-JUN-1999; 99US-0139458.
PR	18-JUN-1999; 99US-0139459.
PR	18-JUN-1999; 99US-0139460.
PR	18-JUN-1999; 99US-0139461.
PR	18-JUN-1999; 99US-0139462.
PR	18-JUN-1999; 99US-0139463.
PR	18-JUN-1999; 99US-0139750.
PR	18-JUN-1999; 99US-0139763.
PR	21-JUN-1999; 99US-0139817.
PR	22-JUN-1999; 99US-0139899.
PR	23-JUN-1999; 99US-0140353.
PR	23-JUN-1999; 99US-0140354.
PR	24-JUN-1999; 99US-0140695.
PR	28-JUN-1999; 99US-0140823.
PR	29-JUN-1999; 99US-0140991.
PR	30-JUN-1999; 99US-0141287.


```
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 21-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 28-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 11-AUG-1999; 99US-0148171.
PR 12-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149368.
PR 18-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 26-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.

PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 65.5%; Score 2572; DB 21; Length 850;
Best Local Similarity 62.0%; Pred. No. 8e-210;
Matches 453; Conservative 118; Mismatches 128; Indels 32; Gaps 6;

Qy 8 LLLVICL-----LDFFSSVKASVSYDDRAIIINGKKILISGSIHPRSTPQMWDLIQ 61
Db 10 lilwclgfililgvf---vcgvtydrkallingqrilfsgsihyprstpdnwdliq 66
Qy 62 KAKDGLDVIETVYFWNGHEPSPCKYNFEGRYDLVRFTKMWORAGLYVNLRIgPYVCAEW 121
Db 67 kagggdvletyvfwnlhepspgkydfegrndlvrvktlnkaglyahnlrigpyvcaew 126
Qy 122 NFGFPVWLKYVPGMEFRNNQPKVAMQGFQKIVNMKSENLFESOGGPIIMAEIENE 181
Db 127 nfggfpwlkyvpgisfrtdnepfkramkgfterivelmksefnifesgggipilisqiene 186
Qy 182 YGPVENEIGAPGKAYTKWAAQMAVGLTGPVIMCKQEDADDPVIDTDCNGFYC6GFRPNK 241
Db 187 ygrggqllgaeghnymtwaaakmatetgvpwmckeddaddpvpintcngfydcfsfapnk 246
Qy 242 PYKPKMTEVWTGMYTKFGGPIQORPAEDIAFSVARFVQNNNGSFNXYMHGTFNGRTS 301
Db 247 pykplwteaawsgwtfegpnmhrpvqdlafgarfikggsfvnyymyhggnfgrta 306
Qy 302 SCLFIATSYDYDAPLDVEYLLNEPKYGHRLDLHKAIKLSEPALYSSVAATSLGSNOE-- 359
Db 307 ggpfttsydydapideyglirqpkyghlkelhrai kmcekalysadvpvtstgnkqgv 366
Qy 360 -----AHVYRSKSGACAAFLSNYDSRYSVKVTQNFONRPNYLPWPWSISILLPCKTAVYNTA 413
Db 367 iyyerfahvysaesgdcsaflnydtesaarvlfnvhnlyppwsisilpdcnrvnta 426
```

Qy	414	QVNSOSSIKMTPAGGSLWSQSYNEETPTADSDTLTANGLEWQKNVTRDSSDYLWYMTN	473	PR	14-MAY-1999;	99US-0134221.
Db	427	kvsn-----fqwesyledssldssfttchllqinvtrdsdylywmts	473	PR	14-MAY-1999;	99US-0134370.
Qy	474	VNIASNEGLKNGKDPYITVMSAGHLVHFVNGKLSGTVYGTLDNPKLTYSGNVKLRAGI	533	PR	14-MAY-1999;	99US-0134768.
Db	474	vdldgsesflhgeltplliidgstghavhifvngqlsgsaftrqrrfyygknlhsgt	533	PR	14-MAY-1999;	99US-0134941.
Qy	534	NKISLLSVGLPNVGVHYDTWAGVLGPTVLTSLGNEGRNLAKQWSYKVLKGESLSL	593	PR	14-MAY-1999;	99US-0135124.
Db	534	nrllalsvavlpnvghfswntgilpvalhslsgqkmdlsqwkcyqvglkgeammI	593	PR	14-MAY-1999;	99US-0135353.
Qy	594	HSLGSSSVVWRSLMAQK-OPLTWYKATFNAPCGNDPLALDMSMGKGQIWIINGCVG	652	PR	14-MAY-1999;	99US-0135629.
Db	594	afptnpsigwmdasltvqkppltwhktyfdapegneplaldmegmgkgiwngeisg	653	PR	14-MAY-1999;	99US-0136021.
Qy	653	RHWPGYIAQDCSKCSYAGTNEKKCQTNCGQPSQRWYHVPRLKPSGNLLVVFEEWGG	712	PR	14-MAY-1999;	99US-0136392.
Db	654	rywtaf-atgdcshcsygtgtykpnkcgctgcgqptqrwyhvprawlkpsqnlvifeelgg	712	PR	14-MAY-1999;	99US-0136782.
Qy	713	NPTGISLVRRS 723		PR	14-MAY-1999;	99US-0137222.
Db	713	npstsvlvkrs 723		PR	14-MAY-1999;	99US-0137528.
RESULT 12						
AAG39970						
ID	AAG39970 standard; Protein; 850 AA.					
AC	AAG39970;					
XX	18-OCT-2000 (first entry)					
DE	Arabidopsis thaliana protein fragment SEQ ID NO: 49530.					
KW	Protein identification; signal transduction pathway; metabolic pathway;					
KW	hybridisation assay; genetic mapping; gene expression control; promoter;					
OS	Arabidopsis thaliana.					
PN	EP1033405-A2.					
PD	06-SEP-2000.					
PF	25-FEB-2000; 2000EP-0301439.					
XX	25-FEB-1999;	99US-0121825.		PR	14-MAY-1999;	99US-0134221.
XX	05-MAR-1999;	99US-0123180.		PR	14-MAY-1999;	99US-0134370.
XX	09-MAR-1999;	99US-0123548.		PR	14-MAY-1999;	99US-0134768.
XX	23-MAR-1999;	99US-0125788.		PR	14-MAY-1999;	99US-0134941.
XX	25-MAR-1999;	99US-0126284.		PR	14-MAY-1999;	99US-0135124.
XX	29-MAR-1999;	99US-0126785.		PR	14-MAY-1999;	99US-0135353.
XX	01-APR-1999;	99US-0127462.		PR	14-MAY-1999;	99US-0135629.
XX	06-APR-1999;	99US-0128234.		PR	14-MAY-1999;	99US-0136021.
XX	08-APR-1999;	99US-0128714.		PR	14-MAY-1999;	99US-0136392.
XX	16-APR-1999;	99US-0129845.		PR	14-MAY-1999;	99US-0136782.
XX	19-APR-1999;	99US-0130077.		PR	14-MAY-1999;	99US-0137222.
XX	21-APR-1999;	99US-0130449.		PR	14-MAY-1999;	99US-0137528.
XX	23-APR-1999;	99US-0130510.		PR	14-MAY-1999;	99US-0137502.
XX	23-APR-1999;	99US-0130891.		PR	14-MAY-1999;	99US-0137724.
XX	28-APR-1999;	99US-0131449.		PR	14-MAY-1999;	99US-0138094.
XX	30-APR-1999;	99US-0132048.		PR	14-MAY-1999;	99US-0138450.
XX	04-MAY-1999;	99US-0132484.		PR	14-MAY-1999;	99US-0138847.
XX	05-MAY-1999;	99US-0132485.		PR	14-MAY-1999;	99US-0139119.
XX	06-MAY-1999;	99US-0132486.		PR	14-MAY-1999;	99US-0139452.
XX	07-MAY-1999;	99US-0132487.		PR	14-MAY-1999;	99US-0139453.
XX	11-MAY-1999;	99US-0132863.		PR	14-MAY-1999;	99US-0139454.
XX	14-MAY-1999;	99US-0134256.		PR	14-MAY-1999;	99US-0139455.
XX	14-MAY-1999;	99US-0134218.		PR	14-MAY-1999;	99US-0139456.
XX	14-MAY-1999;	99US-0134219.		PR	14-MAY-1999;	99US-0139457.

```
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0146389.
PR 04-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 23-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 26-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 10-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.

PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 65.5%; Score 2572; DB 21; Length 850;
Best Local Similarity 62.0%; Pred. No. 8e-210;
Matches 453; Conservative 118; Mismatches 128; Indels 32; Gaps 6;

QY 8 LLLVLC-----LDFFSSKASYSYDDRAIILNKRLILSGSIHYPRSPOMMPDLIO 61
Db 13 LILWfcglflllgvf---vqcgvtcdkallingrillfsgshyprispdmmedliq 69
QY 62 KAKDGLDVIETYYFWNGHEPSPGKYNFEGRYDLVRFIKMVQRAGLYVNLRIQPYVCAEW 121
Db 70 kakdggldvietyvfnlhpepspgkydfegrndlvrfvktihkaglyahlrigpyvcaew 129
QY 122 NFGGFPWLKYVPGMEFRTNNQPKVAMQGFQKIVNNMKSENLEFESOGGPIIMQAIONE 181
Db 130 nfggfpwlkyvpgisfrtdnepfkrankgfteriveimksenlfesggpillsiqene 189
QY 182 YGPVEWEIGAPGKAYTKWAQOMAVGLKTGVPWIMCKQEDADPPVIDTCNGFYCEGFRPNK 241
Db 190 yrggqllgaeghnymtwaakmaiatetgvpwmmckeddapvpvintcngfydcfapnk 249
QY 242 PYKPKMTVEVWTGWYTKFEGPIQORPAEDIAFSVARFYQNNNGSFFNYHYHGGTFNGRTS 301
Db 250 pykpliwtcaewsgwftfeggmhhrpqdlaqfvarfiqkggsfvnyymhgtnfgta 309
QY 302 SGLFIATSYDYDAPLDEYGLLNEPKYGHRLDHLRAIKLSEPALVSSYAAVTSLSGNQBP-- 359
Db 310 ggpvttsydydapideyglirqpkghlkelhrai kmcekalyvadbpvvtisgnkqgvw 369
QY 360 -----AHVYRSKSGACAFLSNYDSRYSVKVTQFONRPNYLPWPMSISILPCKTAVYN 413
Db 370 iyyerfahvysaesgdsafianydtesaarvlfnnvhyhnpwpwsisilpdcrnafinta 429
QY 414 QVNSQSSSIKMTAPAGGSLWSQSYNEETPTADDSDTLTANGLEQKNVTRDSDYLWYMTN 473
Db 430 kvsn-----fqwesyledsslddsstftchgleqinvtrdtsdylwmts 476
QY 474 VNIASNEGLKNGKDPYLTVMASAGHLVHFVNGKLSGTVYGTLDNPKLTYSGNKLKLAGI 533
Db 477 vdigdsesflhggepltiidstghavhifvngqlsgsafgrqrfrfytgqkinlhesgt 536
QY 534 NKISLLSVSGLPNVGVHYDTWNAAGVLGPVTLISGLNEGRNLAKOKWSYKVKLGESLSL 593
Db 537 nriallsavglpnvgghfswntgilgpvalhglsgqmdlsqwkqcyvgylgkeamnl 596
QY 594 HSLSGSSVVEVRGSLMAQK-QPLTWYKATFNAPGNNDPLALDMSMGKGQIWIINGEGVG 652
Db 597 afptnptsigwmdasaltvqkqppltwkhtyfdapegneplaldmegmkgkglwngesig 656
QY 653 RHWPGYIAQGDSCSKSYAGTNEKKCQTNCGQSPQSORWYHVPKSLKPSGNLLVFEPEGG 712
Db 657 rywtaf-atgcdschsytytkpnkcgqgqptqrwyhvpawlkpsqnlvlvifeelgg 715
QY 713 NPTGISLVRRS 723
Db 716 npstsvlvkrs 726

RESULT 13
AAG31160
ID AAG31160 standard; Protein: 853 AA.
XX
AC AAG31160;
XX
```



```
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 63.3%; Score 2482.5; DB 21; Length 794;
Best Local Similarity 64.1%; Pred. No. 3e-202;
Matches 430; Conservative 110; Mismatches 128; Indels 3; Gaps 3;

Qy 55 WPDLLIQAKDGLDVTETVFNHNGHPSPKYNPEGRYDLVRKMKVQVORAGLVNLRIG 114
Db 1 mmedliqakdgddvdietyvfnlhhepsdkdyfegndlvrvfktihkaglyahlrig 60

Qy 115 PYVCAENFGGPPVWLKYPVGMERTNNOPEKVMQGVQKIVNMKSENLFESGGPII 174
Db 61 pyvcaenfggppvwlkypvgsirtdnephframkgfterivelmkseinfesgggpii 120

Qy 175 MAQIENEYGPVEWETGPKRAYTKWAQMAVGLKTGVPWIMCKQEDADPDVIDTCNGFYC 234
Db 121 lsqiencygrgqqlgaeghymtwaakmaiatetgvpwmckeddapdpvintengfyc 180

Qy 235 EGFPRNPKYPKMMTEVWNTGYTKFGGPIQORPAEDIAFSAVFQNNQSFNMYHGG 294
Db 181 dsfapnkpypkpliwteaawgftfeggmhbrpvgdlaafvarfiqkgstfvnyymhyg 240

Qy 295 TNFGRTSSGLFIANSYDAPLDVEGLLNEPKYGHRLDKAIKLSPEALVSSAAVNTSL 354
Db 241 tnfgrtsgglfiansydydapldveglgllrtpkyghlkeihraikmcekalvsadpvtlsi 300

Qy 355 GSNQEAHYRSKSGACAFSLSNYDSRYSKVYTFONRPNLPPWSISILPDCKTAVYNTAQ 414
Db 301 gnkqahvysaesdcasflanydtesaarvlfnnvhnlpwsisilpdcnnavfntak 360

Qy 415 VNSQSSSTKWTAG-GLLSWSYNEETPTADDSDTLTANGLWEOKNVTDRSDSLWYMTN 473
Db 361 vgvqtsgmemlptdtknfwesyledissldsdstftthgleginvtrtdsdywmynts 420

Qy 474 VNIAENEGFLKNGDPYLTVMAGHLVHFVNGKLSGPFVYCTLDNPKLTYSGNVKLRAGI 533
Db 421 vdiqdesflhgaelptliqstghavhifvngqlsgsafqrurfrtygqknlhngt 480

Qy 534 NKISLLSVGLPNVGHYDVTWAGVLGPVTLISLINEGRNLAKOKWSYKVKLGESLSL 593
Db 481 nriallsavglpnvgghfswntgilgpvalhlsqgkmdlswqkwtqyqglkgeamnl 540

Qy 594 HSLSGSSVENVRGSLSMAQK-OPITWTWKATENAPGNDPLDALDMASMGKGQIWIENGEGV 652
Db 541 afptntpsigwmdasltvqkbpqitwhktyfdapegneplaldmegmkgqilwvngesig 600

Qy 653 RHPWFIYAQGDGCSYAGTNEKCKQNCQPSQWRVHVPKSLKPSGNLLVYFEEWGG 712
Db 601 rywtai-atgdcshcsytgtkpnkcqcgqgptqrwyhvprawlkpsqnlvifeelig 659

Qy 713 NPTGISLVRRS 723
Db 660 npstsvlvkrs 670
```

RESULT 15

AAG39972
ID AAG39972 standard; Protein: 788 AA.
XX
AC AAG39972;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 49532.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EPI033405-A2.
XX
PD 06-SEP-2000.
XX
25-FEB-2000; 2000EP-0301439.
XX
25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.

PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 24-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.

PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161902.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match Best Local Similarity 62.3%; Score 2446.5; DB 21; Length 788;

Matches 426; Conservative 109; Mismatches 120; Indels 23; Gaps 4;

Qy 55 MWPDLIQAKDGLDVIETVYFWNGHSPSGKYNFEGRYDLVRFTKMVORAGLYVNLRIg 114
Db 1 mwedliqakdggidvetyvwnlhhepspgkydfegndlvrfvktlnkaglyahlrig 60
Qy 115 PYVCAEWNEFGGFPWLYKVPVGMERTNNQPKVAMQGVQKIVNMKSENLFESOGGPII 174
Db 61 pyvcaewnfpgfpwlykvpvgsifrdnephframkgfiterivelmksenfesqggpII 120
Qy 175 MAQIENEYGPVWEIGAPKAYTKWAQMAVGLKTGVPMCKQEDADPDVIDTCNGFYC 234
Db 121 lsqieneygrqgqlgaeghnytmwaakmaiatetgvpwvnmckeddapdpvintcngfyc 180
Qy 235 EGFRPNKPKPKMTEVWTGWYTKFGGPIPORPAEDIAFVARFVNNGSFNFYMYHGG 294

[illegible]

Search completed: June 24, 2002, 20:38:20
Job time: 155 sec

THIS PAGE BLANK (USPTO)